The Perception of Surface Properties: Translucence and Gloss

CHADWICK, ALICE,CAITLIN

The Perception of Surface Properties: Translucence and Gloss

A thesis submitted for the degree of Doctor of Philosophy in the Department of Psychology

University of Durham

2016

Alice Caitlin Chadwick[1]

[1]Studentship jointly funded by the Department of Psychology, University of Durham, and Procter & Gamble, Technical Centres, Newcastle UK (grant number RF010063).

Abstract

The human visual system is sensitive to differences in gloss and translucence, two optical properties which are found in conjunction in many natural materials. They are driven by similar underlying physical properties of light transport - the degree to which light is scattered from the surface of a material, or within the material. This thesis aimed to address some fundamental questions about how gloss and translucence are perceived. Two psychophysical methods (maximum likelihood difference scaling, and conjoint measurement) were used throughout, as they provided an appropriate way of investigating how perceptual experiences related to physical variables.

In the introduction, I review the literature on the perception of gloss and translucence. Study 1 investigated the relationship between variables controlling light transport in translucent volumes and percepts of translucence. The results show that translucence perception is not based on estimates of light transport properties per se, but probably uses spatially-related statistical pseudocues in conjunction with other cues. Study 2 examined a similar issue, but the translucent material was presented as a layer enveloping a solid object. Behavioural responses were similar for these translucent materials, which were perceived as glossy layers of coating. Study 3 further explored established findings that perceived translucence shows inconstancy under changes in viewing condition. Perceived translucence was dependent in a complex way on both light-scattering in the material and illumination direction in both volumes and layers of translucent materials. Study 4 used similar layers of subsurface light-scattering and -absorbing material and applied them to multiple base materials. Opacity and a lack of mirror-like reflections enabled observers to make the most accurate independent judgements of darkness and cloudiness.

Study 5 explored observers' sensitivity to spatial variation of scatter across a surface using similar layers of coating, and the way in which observers might weight cues dif-

ferently to answer subtly different questions (judgements of 'shininess' vs. 'cleanliness'). Layer thickness and variation of scatter significantly affected perceived shine and cleanliness, with layer thickness influencing decisions more than variation. Scatter variation contributed to decisions significantly more for judgements of cleanliness than shine. Study 6 investigated how tactile surface roughness influenced perceived gloss. Previous findings have shown that tactile compliance and friction influence perceived gloss, and that friction interacts with visual gloss. Our results showed that surface roughness and visual gloss both affected perceived gloss, but there was no interaction, suggesting that different types of haptic information are combined with visual information differently. Finally, study 7 explored the potential cortical basis of perceived translucence. Through testing a neuropsychological patient, we showed that perceived translucence is dependent on cortical areas not responsible for colour or texture discrimination.

The thesis concludes with a discussion of additional recent findings, the implications of the research reported in this thesis, and proposals for future research.

Declaration

I declare that no part of the material presented in this thesis has previously been submitted for a degree in this or in any other University. Where material has been generated through collaboration, this has been indicated where appropriate, and the contributions of the authors made explicit. In all other cases, the work of others has been acknowledged and referenced appropriately.

Published and submitted work, and formats of other chapters

The entirety of Chapter 1.5 is published in Vision Research:

Chadwick, A. C., & Kentridge, R. W. (2015). The perception of gloss: a review. *Vision research, 109,* 221-235.

This paper was researched and written entirely by me, and was edited by my supervisor, the second author. Parts of this chapter were also presented at the Pigment & Colour Science Forum and Ti02 2015, Prague:

Pigment and Colour Science Forum and Ti02 (2015). *The Perception of Gloss: A Review. A.C. Chadwick, University of Durham.*

The entirety of Chapter 2 is under submission with Current Biology, with authors A. C. Chadwick, G. Cox, H. E. Smithson, and R. W. Kentridge. The contributions of each are as follows: the initial hypotheses were formulated by H.E.S. and R.W.K., the experiment was designed by H.E.S., R.W.K. and A.C.C., G.C. collected data with the real stimuli, A.C.C. collected data for rendered stimuli and conducted MLCM and MLDS analyses for all data, R.W.K. created the rendered stimuli, and G.C. created the real stimuli under the supervision of H.E.S. The paper was written by A.C.C., H.E.S., and R.W.K., and H.E.S.

calculated image statistics and model fits for simulated tasks. As such, it is also formatted in the style required by this journal. Some minor amendments have been included here, for purposes of clarity within the thesis.

The entirety of Chapter 8 is currently under submission with Cortex, with authors A. C. Chadwick, C. A. Heywood, H. E. Smithson and R. W. Kentridge. The contributions of each are as follows: The experiments were designed and conducted by A.C.C. and R.W.K, and A.C.C. analysed the data and wrote the paper. C.A.H. provided assistance with testing M.S. on the Farnsworth-Munsell 100 Hue Test and edited and gave feedback on drafts of the manuscript. H.E.S. calculated image statistics of the images, and provided assistance with modeling the image statistics of best fit for the obtained data. This paper is formatted in the style required by this journal. Referencing has been altered for consistency with the rest of the thesis.

Where chapters have not yet been submitted to journals, they are presented in a standard APA format.

Acknowledgements

First, I thank Procter and Gamble for providing me with this studentship, and the wonderful opportunity to work on this project.

I am extremely grateful to my primary supervisor, Bob Kentridge, for all of his support and guidance, and for putting up with my Gantt charts. It has been a privilege and a pleasure to work with him, not to mention a very steep learning curve. Thanks also to Charles Heywood for his kindness and support, and who, together with Alex Easton set up this project and acted as secondary supervisors. I am incredibly grateful to Hannah Smithson for collaborating with us so extensively, and without whose tutoring in colour vision and encouragement I would not have dreamt of applying for a PhD.

I must also thank all of the stellar people who so generously participated in my experiments, which were often not the most interesting of tasks. Particular thanks to the various inhabitants of Room 70, for advice, pub trips, and for making me feel so welcome, in spite of often being cajoled into being my pilot guinea pigs - including Emma L, Emma G, Hannah H, Hannah R, Yan, Pete, Kayleigh, and Dave. Thanks also to Liam for all of his advice and frequent last-minute participation in studies, and to the more recent arrivals - Ali, Tom and Emily - for putting up with my writing-up cabin fever.

I'm particularly grateful to all of my wonderful Ustinov friends as well; you made life outside the PhD a great deal more enjoyable, and I feel very lucky to have met such a fantastic group of people from all around the world.

Most of all, I am thankful to my parents, for their unwavering support, encouragement, and kindness. This thesis could not have been done without them. Thank you also to Patrick, for his never-ending support, and for having so much confidence in my abilities - that makes one of us at least.

This thesis is dedicated to my mum and dad.

Contents

List of Tables

List of Figures

Chapter 1

General Introduction

1.1 How do we perceive materials?

How does the visual system represent the properties of materials? It seems intuitive to us that objects have shape, colour and texture, but we are still uncertain whether these properties are represented separately in the visual system and exactly how perceptual judgements are made.

Generally, while human vision seems to us to be rich, effortless, and almost infallible, in fact our visual experience is based on complex extraction of information from a limited sensory input. The light that reaches the eye is already a scrambled combination of illuminations and reflectance profiles, which cannot be easily untangled: the illuminant in any scene (e.g. skylight, direct sunlight, or artificial illumination) is composed of different amounts of each wavelength of light, and all of the surfaces in a scene have individual reflectance profiles, reflecting different amounts of light at each wavelength. The illumination is therefore absorbed or reflected - according to the individual surface reflectance profiles - by the surfaces in the scene, significantly altering the spectral composition of light that subsequently reaches an observer. This complex combination of information is detected by just three types of cone receptor, each type optimally tuned to respond to a particular range of wavelengths, essentially limiting the sensory input to three cone signals, which can also be noisy. It then seems an almost impossible task for the brain to pick apart these signals and piece together colours, lightnesses, and borders, coherent shapes and objects, and properties of materials from this restricted information. There are consequently a number of contentious problems in the field of visual perception, such as colour constancy, scene parsing, and the integration of information. It is not for this thesis to explain how the visual system achieves all this. However, some of the more general theories of perception are relevant and indeed of primary importance to consider here.

Two prominent theories of how the visual system interprets information are the theories of inverse optics and of short-cut statistics, or simple image heuristics. The theory of inverse optics proposes that the brain is able to reverse-calculate surface reflectances, illuminations, and therefore properties of objects and scenes around us (see Pizlo 2001). This implies that the brain 'knows' the physical equations governing the interaction of light with materials, and can therefore make accurate estimates of illumination compositions and surface reflectance characteristics, and 'unpick' properties from the three types of cone signal. In contrast, the theory of short-cut statistics proposes that it is impossible for the brain to reverse-calculate these properties, and so instead it calculates simple statistics from the signals received and uses these to approximate - or act as a proxy for - changes in physical parameters, with different simple statistics acting for different properties or characteristics (see Kersten 2000, and Motoyoshi, Nishida, Sharan, and Adelson 2007).

However, both of these theories appear to be seriously flawed, and are indeed polarised at opposite ends of the range of possible solutions. An inverse optics approach has failed on numerous occasions when studying problems of vision - in particular, colour constancy. It is impossible for the visual system to get an accurate estimate of the spectral composition of the illuminant, by any of the methods proposed; and it is impossible to disentangle this from the complex combination of light that reaches the retina. A short-cut simple statistical approach - while avoiding the issues of the inverse optics view - also fails to take into account the full range of information important for making visual judgements. Short-cut statistics have been posited as the cues used by observers to estimate many visual properties such as gloss, shadows, shape, and depth. However, more often than not, the simple statistics proposed do not correspond with the way in which human observers make judgements. For instance, Motoyoshi et al. (2007) proposed skew of luminance as a robust cue for perceived gloss, but on further inspection this cue does not take into account many kinds of other information already confirmed as being important for observers when making decisions of glossiness (see Landy 2007) - in particular, the spatial relation between specularities and the three-dimensional shape of the object, shading, texture, and lighting direction. In addition, Landy showed that manipulation of the one simple statistic alone only manipulated perceived gloss in a restricted set of stimuli, and in fact was a poor predictor of perceived gloss in other images. Motoyoshi (2010) proposed a simple statistic that might be used as a cue for predicting perceived translucence and transparency, but again there was a restriction on the range of materials and objects to which the cue applied. While each approach is useful as a method of investigation, clearly neither can be the one correct answer. Indeed, the most extreme version of an 'inverse optics' model where the brain knows the physical equations describing the transport of light and its interaction with materials and can reverse-calculate the actual properties of the scene is now regarded as more of a 'straw man', due to the fundamental ambiguity between the information on the retina and the true characteristics of the source of this information in the scene.

Current consensus in the literature generally seems to converge on a middle ground between these polar opposites. There is evidently something clever about information interpretation achieved by the visual system, and it now seems likely that more complex spatially constrained image statistics - weighted differently according to context - might be the best approximation we can achieve. Many different types of information are important in making visual judgements, but variations within and between individuals imply that these are not used consistently or accorded the same priority. These different kinds of information are referred to as 'cues' in the literature. I argue that the specific term 'cues' should be used to refer to the physical aspect of these properties, and that 'pseudocues' should be adopted to refer to the information which is in practice used by observers to infer the properties of surfaces and objects when making decisions. Cues are not the same as object properties, as object properties are not directly available to the observer - cues are image properties which covary more directly with object properties than pseudocues; for example, the ratio of brightness in specularity and brightness in the surrounding scatter might be considered a cue for gloss, as gloss is the ratio of specular and diffuse reflectance. Other things like distinctness-of-image in a surface reflection are most certainly pseudocues: pseudocues are not independent of object properties, but rather are indirectly affected, as with distinctness-of-image in surface reflections. The physical ratio of specular to diffuse reflectance from a surface does not directly affect the distinctness-of-image on a surface, but the smoother and more mirror-like a surface, the more it will reflect a higher ratio of specular to diffuse light, and the higher the clarity of reflected images. Furthermore, glossy surfaces look darker than their matte counterparts, as contrast increases between the highlights and diffuse background, so this contrast may further enhance the perceived distinctness-of-image. Gloss could not be inferred from simple mathematical transformations, as the relationship between physical gloss and a pseudocue is complex, but if an observer had a number of pseudocues despite the fact that the relationship was complex, then they can make a more reliable inference. For example, a change in distinctness-of-image might not be detectable with a particular change in physical gloss for lots of images, but some other pseudocue might change and this could be used to infer the difference. For instance, contrast gloss is a pseudocue as well: increasing surface gloss makes materials appear darker as the contrast between highlight and diffuse background increases, however, if darkness is manipulated while surface gloss is held constant, the surface will also appear more glossy (Beck, 1964; Harrison & Poulter, 1951). Pseudocues are not necessarily related directly to the surface properties themselves, but may cause similar perceptual changes.

Gloss might not not always affect image properties monotonically. In contrast, image statistics could mean one of two things - image statistics could theoretically comprise a full description of an image, or it could be a very simple statistic such as Motoyoshi's luminance histogram, which is insensitive to locations within the image. For the purposes of this thesis, I will describe the latter as simple image statistics'. Simple and complex image statistics have both been proposed as possible pseudocues, but simple statistics such as Motoyoshi's are not plausible as pseudocues as they are not sufficiently informative or generalisable. More complex image statistics are more likely to be used as pseudocues, as these are capable of being location dependent and dependent on object geometry. Image statistics calculated in later chapters are of both the simple and complex variety, and will be indicated as such.

'Cues' would therefore relate more closely to the inverse optics end of the spectrum, and 'pseudocues' appear in-between the two polar opposites of inverse optics and short-cut simple statistics, potentially involving complex weighted spatially-related image statistics. Among those supporting such a middle-ground theory of perception is Fleming (2014), who concludes that the visual system seems to parse scenes into particular appearance characteristics, employing a generative statistical appearance model with which to infer properties that vary between samples. Such a model does not try to estimate the underlying physical properties of surfaces, or the composition of light reflected back to the observer. Furthermore, this approach offers a clear advantage over simple image heuristics, in that it extends to judgements of unfamiliar materials.

A generative model such as Fleming's is different from an inverse optics model in that the observer is trying to estimate or approximate the generative variables such as shape, reflectance and illumination based on an internal model of their joint effects on the image, rather than the actual physical characteristics of the scene. A generative model is a more plausible alternative to both an inverse optics model and to a short-cut simple statistical model, as it is both more achievable and more likely to be generalisable to other stimuli. A generative model of gloss would involve an internal model where gloss is the ratio of specular to diffuse reflection, and estimates for the parameters of this generative model would be produced. The idea of pseudocues does not conflict with a generative model. It is plausible that pseudocues or cues could be the method by which an observer estimates the parameters in the generation of gloss, where the pseudocue has been identified as something that covaries with the relevant variables in the generative model, however one could imagine another approach where an observer purely responds to pseudocues. That is, the presence of a pseudocue produces a percept, but there would not be an underlying internal generative model of optics - so pseudocues could fit into both a model used to infer the ratio of specular to diffuse reflectance, or they may be arbitrary heuristics. It is also possible that, within the context of a generative model, if an observer uses pseudocues they may ignore generative variables other than the one being estimated. A pseudocue would be used if it works well enough to provide a sufficient estimation, and it could potentially be used as a proxy for all possible generative variables - indeed, our perceptual constancy for many judgements is not perfect, so this too is plausible.

1.2 Gloss and translucence

Much previous research on perceived surface properties has focused on colour vision, and a great deal is now known about the way in which we perceive colour - ranging from genetic influences through physiological and cortical representation to cognitive factors. However, less is known about how we perceive other properties such as gloss (or shine) and translucence (please note - I will be using the terms ‘gloss’ and ‘shine’ interchangeably). Transparency is the physical property of allowing light through a material without being scattered but simply refracting at the edges of the material where changes in density occur. Translucence is a superset of transparency, where light is allowed to pass through a material and is refracted by the material, but light can be scattered internally. Opacity does not allow light to pass through a material at all, and transmits no light - all light is absorbed, scattered, or reflected. The physical definition of gloss at a surface is the ratio of specular to diffuse reflection, and surface roughness is the polar opposite of this ratio scale (where there is more light diffusely reflected at various spatial bandwidths than specular reflection). These definitions are the same as those for optics. Figure 1.1 illustrates light transport for gloss and for translucence in a thin layer, to further clarify the difference between these two concepts. As will be further explained in this thesis, variations in gloss of different materials can be approximated by manipulating a very thin layer of a translucent scattering material. The study of the perception of surface texture has developed separately from the study of colour, as it was initially assumed that surface texture was processed in parallel with colour; however recent studies have suggested that this is not necessarily the case (Cavina-Pratesi, Kentridge, Heywood, & Milner, 2010a, 2010b). In addition to the study of texture perception, translucence and gloss are two properties that have also been neglected to date. As Viénot reports, the study of perceived gloss is an interesting field because of the ways in which gloss can be interpreted. Gloss is a physical description of a material, related to physical properties, but perceived gloss is related to human vision; an ‘appearance attribute appraised privately, dependent on cognitive cues and other appearance attributes’ (Viénot, 2012). There is an inherent difficulty in the study of perceived gloss, in determining how the perceptual and physical interpretations can be linked. In the second part of this introduction, I explore and summarise the field of gloss perception from a historical perspective, and present the findings and overall conclusions to date (at the time of its publication).

In real world and particularly natural materials, gloss is often accompanied by a certain level of translucence - either in a bulk volume, or in layers of translucent coating, or be-

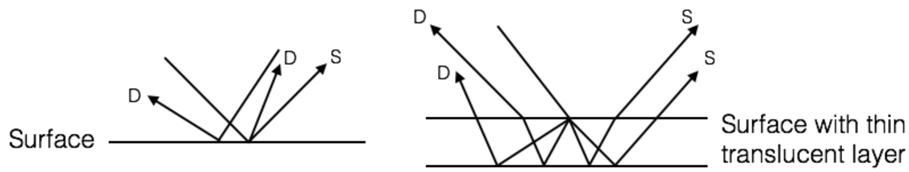

Figure 1.1: Differences in light transport for opaque surfaces and layers of translucent material. 'D' indicates diffusely reflected light, and 'S' indicates specularly reflected light.

neath the surface of materials such as marble, fruit flesh, and skin. Often glossy materials do not just reflect light from the surface; some light travels beneath the outer layers, even if only a short distance below the surface. Human observers are sensitive to these light volume transport effects (Koenderink & van Doorn, 1980), but they are poorly understood. Two key factors determine volume transport - light scatter and absorption. This prompts a question: do observers make judgements about translucent materials by estimating these light transport properties, or by making estimates of the material using pseudocues such as complex image statistics? The degree of scattering influences perceived glossiness whether on the surface or in sub-surface volume transport (see Chapter 6). Changes in perceived gloss are often driven by changes in the scattering of light in these layers of translucent materials, so the study of gloss should involve an investigation of the properties of these translucent layers. To date, there is little experimental research into perceived translucence. Previous studies have either proposed an image-heuristics approach, or proposed a middle-ground approach, in neither case investigating judgements from observers; I will now report these findings.

1.3 The history of research in translucence

Research into perceived translucence began fairly recently in comparison with work on other visual properties such as colour. Metelli (1970), while not the first to study the phenomenon, made one of the first significant attempts to explain perceived transparency. He outlined potential algebraic conditions under which we perceive transparency, in terms of relative figural relationships and ordinal relationships between colours in a scene. Metelli, however, was defining such relationships in very limited cases where the transparent substance was a fine film covering a boundary between two different colours. Furthermore, conditions were in relation to criteria defining where transparent objects might be interpreted, rather than exploring how observers perceive the translucence of the object in practice. In subsequent work Metelli proposed that perceived translucence was the result of splitting the stimulus luminance between the luminance of the background and that of the transparent surface (Metelli, 1974a, 1974b). He also identified three main figural conditions for perceiving transparency (figural unity of the transparent layer, continuity of boundaries in the transparent region with boundaries of the non-transparent region, and stratification of the transparent region into two overlapping layers). Again, these proposals were more concerned with defining conditions that might have to be satisfied for the visual system to classify a surface as transparent, rather than explaining how we perceive the translucence itself. Similarly, Brill (1984) developed a form of inverse-optics diagnostic for heuristics that could be used to identify a translucent material through a set of physical constraints. However this was limited to a specific kind of translucent materials and certain contexts and did not address the question of how observers perceive the material itself. Much of the research relating to perceived translucence until the early 2000s focused on ways of interpreting translucence in terms of relations between surfaces or descriptive laws (Adelson & Anandan, 1990; Beck, Prazdny, & Ivry, 1984; Singh & Hoffman, 1998). As technology developed, a model of translucent materials in terms of subsurface light transport was proposed (intended for use with a ray tracing computer graphics image renderer - Jensen, Marschner, Levoy, and Hanrahan 2001). Theories of translucence developed further, with research showing that there were many other unconsidered but important aspects of translucence. For example, shape-from-shading becomes more complex for translucent objects with complex geometries; contrast was shown to be more appropriate as an image variable in assigning transmittance to transparent differences (rather than luminance differences, as Metelli had predicted); and blurring in a scene was concluded to contribute towards perceived translucence, but was not sufficient to explain decreases in perceived translucence in terms of contrast alone (Koenderink & van Doorn, 2001; Singh and Anderson 2002a; Singh and Anderson 2002b).

Until this point, most published work discussing the perception of translucence had not attempted to address the fundamental questions of how it was perceived. It was evident that observers were very good at discriminating and identifying differences in material properties including opacity and translucence, even with very low presentation times (Sharan, Rosenholtz, & Adelson, 2009). Fleming and colleagues concluded that many of the simple cues for perceived translucence proposed were, so far, insufficient and unable to predict how the translucence of a material was perceived (Fleming & Bülthoff, 2005; Fleming, Torralba, & Adelson, 2004). It was argued that the physics of translucency - much like the physics of surface reflectance, in colour vision - were too complex for the visual system to be able to estimate using inverse optics. In addition, the authors argued that simple image statistics were inadequate when used alone, and while the relevant sources of information were unknown they proposed that perceived translucence was achieved by means of parsing scenes into key regions and gathering image statistics from those regions. Some research persisted in examining simple image statistics as solutions for the problem of perceived translucence. However like previous attempts at characterising perceived gloss, the proposed statistics only applied to the particular objects used in the experiment and did not work for all translucent or transparent materials, or in alternate contexts (Motoyoshi, 2010). As the field of material perception expanded, Anderson reiterated the importance of perceived translucence for observers despite the lack of knowledge about how the visual system infers translucency. He agreed that the ability to distinguish material properties on the basis of translucence meant there is information available to observers about the way that materials transport and reflect light (Anderson, 2011).

More recently, Fleming, Jäkel, and Maloney (2011) recognised that previous work had focused primarily on theoretical approaches to perceived translucence, and also on thin filters rather than translucent materials found in volume. Thick transparent objects, irregular in shape and varying in refractive index were considered by the authors. It was concluded that an important part of the evidence indicating the presence of such objects was distortions in perceived shape of other objects present, and that a new class of visual cues derived from these distortion fields could be employed by the visual system. Gkioulekas et al. (2013) made further progress by identifying that multiple scattering (forward and backward scattering of light within a volume) contributed to the translucent appearance of materials. They also showed that the phase function of multiple scattering can contribute to a translucent appearance. This was an important step in being able to characterise the physical characteristics of translucent materials. However, while this manipulation affected perceived translucence it is not necessarily a measure which the visual system attempts to approximate when making judgements of translucence. Beyond pure physical characteristics, it was also found that contextual variation - lighting direction - had a significant effect on perceived translucence. Back-lighting made translucent objects appear more opaque, and front-lighting made them appear more translucent (Xiao et al., 2014).

Research into perceived translucence has evidently been limited, and relatively little is understood about how exactly the visual system interprets translucence. Much like perceived gloss, a consensus appears to have been reached on the kind of information that the visual system is likely to be using; however the precise nature of this information remains unknown. It therefore seems important to address this question alongside the role of scattering of light within volumes in relation to perceived gloss, as it is apparent that both are commonly found together in real life settings.

1.4 The direction of this thesis

There is some debate whether gloss and translucence are processed within the same framework as surface texture. Both result from fine scale variation in the shape of surfaces, but it is not clear whether they are processed in the same way.

The focus of this thesis is to investigate how gloss, and translucence in relation to gloss, are perceived. This is approached at a number of levels. I start by considering the kind of information that observers might be using to make decisions about translucence and gloss. I then explore how this information is affected by other variables in the environment, and whether the perception of gloss and translucence are processed independently from texture. The perception of light scatter is investigated both in volumes, in terms of translucence, and also in the perception of gloss, where surface layer scattering variations are used to apply identical manipulations of physical gloss to different materials and in different contexts. This technique is novel and important as it means that identical manipulations can be performed on a variety of different stimuli.

The first four experiments focus on the ways in which translucent materials (in volumes, and subsequently layers) are perceived. As we do not observe the materials in a single context, the findings from these experiments are then applied in different contexts to assess how perceptions of translucent materials change in relation to lighting direction, and the base material to which the translucent layer is applied.

The following experiment examines how these layers, and variations within them, might influence perception of shine. This study also allows us (with an additional experimental condition) to explore whether decisions based on the perception of a material property such as shine go beyond estimation of physical properties and can be influenced by more abstract knowledge about the concept of cleanliness. The idea behind this study was prompted by the observation that in many cases, objects or surfaces judged as shiny might also be called clean. This question sat within our proposal for a model which shows the many stages at which interactions (of light transport, material properties, cues, and observers' perceptions and judgement criteria) may alter perceptual judgements.

Beyond questions of the fundamental perceptual experience of gloss and translucence, an area currently being explored is the potential integration of perceived gloss with other sensory information. Adams, Kerrigan, and Graf (2016) reported that simulated friction interacted with judgements of perceived gloss. The sixth experiment aimed to investigate how visual information might integrate with tactile feedback from real objects varying in fine-scale roughness, by asking observers to make judgements of shine from incongruent visual and tactile information, presented using a mirror and screen apparatus.

A previous study by Kentridge, Thomson and Heywood (2012) showed that the processing of gloss was not dependent on areas generally thought to be responsible for texture perception, by testing a neurological patient with previously established cortical deficiencies in these areas. The seventh experiment aimed to determine whether the same could be said for aspects of perceived translucence. Through testing the same neurological patient, potential cortical areas responsible for perceived translucence were assessed to see whether these too were independent of the processing of texture.

The second half of this Introduction, and experiments reported in Chapters 2-8, provide a summary of current findings and characterise some essential features of how translucence is perceived and how variation in scattering is involved in the perception of gloss. They also clarify how additional factors such as light direction, base material and spatial variation affect perceived gloss and translucence. The literature review outlines the many factors important in determining how we perceive gloss, but further research is required into the interaction of these factors. The experiments on light direction, base materials, variable scattering in layers, and tactile input begin to address some of the issues affecting our perceptual experience of gloss, but many questions still need to be answered. In the discussion, I propose a multidimensional scaling experiment which would begin to address this question. The design of this experiment is ambitious and technically challenging, and was not feasible within the boundaries of this PhD. As a whole, this thesis aims to begin to characterise the range of factors which influence perceived gloss and translucence, and to show that study of perceived gloss may need to be conducted in conjunction with the study of perceived translucence. In addition, it aims to show the power of new psychophysical methods (maximum likelihood difference scaling and conjoint measurement), and also how these methods - employed jointly - might be better used in future.

Throughout the experimental chapters in this thesis, familiar objects with specific, known materials are used to study the perception of material properties. This approach was used with the intention that different observers would be more likely to interpret the questions we ask of them consistently, as they were already familiar with the ways in which those materials vary and the words with which those changes could be described. There is the possibility that the use of familiar objects means observers may make assumptions about the compositions of those familiar materials and how they behave, which could influence responding (if they alter their responses depending on how they think the material ought to have changed rather than what they are seeing). However, it is hoped that by ensuring the tasks are somewhat challenging, but still achievable, observers remain attentive to the changes visible in the stimuli. The generalisability from findings obtained with familiar objects to novel objects can be debated, yet observers seem to apply prior knowledge of previously encountered objects to novel objects in order to try and make sense of what is in front of them (Burgess, 1985; Kersten, Mamassian, & Yuille, 2004; Krol, 2011; Vandenbroucke, Fahrenfort, Meuwese, Scholte, & Lamme, 2016). They might not be as confident in making judgements of novel objects when applying existing knowledge, so may be more cautious or conservative in their responses, but these findings would still be applicable as observers would be very likely using similar pseudocues and strategies to make their judgements.

As this thesis is in a publication format, with several chapters either already published or under submission to academic journals, I am aware that there is some repetition in places.

1.5 The perception of gloss: a review

Gloss is a relatively little studied visual property of objects' surfaces. The earliest recorded scientific reference to gloss appears to have been by Ingersoll in 1921: studies at this time were based on the assumption that gloss could be understood as an inherent physical property of a surface, and the priority was to devise a satisfactory method and scale to measure it reliably. As awareness of the complexity of perception grew, efforts were made to distinguish different types of gloss, although these generally still took the form of a search for objective physical measures to be solved within the visual system by means of inverse optics. It became more widely recognised approximately 20 years ago that models of gloss perception based on inverse optics were intractable and failed to explain experimental findings adequately. A temporary decline in the number of published studies followed; however the last decade or so has seen a renewal of interest in the perception of gloss, in an effort to map what is now understood to be a complex interaction of variables including illumination, surface properties and observer. This appears to have been driven by a number of factors, as the study of gloss re-emerged from research into other surface properties such as colour and texture, with technological advances paving the way for new experimental techniques and measurements. This review describes the main strands of research, tracking the changes in approach and theory which have triggered new avenues of research, to the current state of knowledge.

1.5.1 Overview

The history of the study of gloss falls into a number of distinct phases: initially, the focus was on finding an objective measure by which materials and surfaces could be compared for physical gloss. Emphasis then shifted to the perceptual aspect of gloss following the work of Hunter (1937), with the recognition that it was more complex than a single physical measure could quantify. For a time continuing research persisted with the theory of a single objective measure of gloss that would supposedly be computed by the visual system using an inverse optics approach. However, the view steadily gained ground that multiple factors must be involved. Work by those such as Sève (1993) underlined the multidimensionality of gloss; the impossibility of obtaining satisfactory measurements using a single instrument to correlate with perceptual judgements; the intractability of an inverse optics approach; and the need for consistent terminology. Focus shifted to the consideration of multiple dimensions of gloss, and the relation between physical and perceptual scales. At the same time there was a separate proposal that the visual system made use of a statistical diagnostic solution, based on a single measurement of regularities in image statistics. However this was not supported and a consensus emerged that a multiple-dimension approach to perceptual gloss was most consistent with the full range of experimental findings. Rather than the visual system attempting to solve inverse optics, or trying to approximate physical dimensions by generalising statistical regularities in a scene, the system treats the multiple dimensions and features within the image as a whole, a gestalt, which leads to a perceptual judgement of glossiness.

1.5.2 Gloss as a single objective measurement

The earliest studies of gloss took it to be a single physical attribute and focused on how to measure it objectively. Ingersoll conducted one of the first studies, examining the measurement of gloss on paper with the use of a glarimeter (Ingersoll, 1921, see Figure 1a). Assuming that gloss could be entirely defined as the amount of specular reflectance of light compared to the amount of diffusely reflected light, the instrument calculated this proportion using a polarising filter (since specularly reflected light had been found to be almost completely polarised). This instrument was put into use in paper mills, in order to determine the quality of the paper produced. Pfund (1930) set out on a similar task, again proposing to measure the specular reflection of various materials. It was a general assumption at this time - and even for the next few decades - that a single objective index of gloss existed, that could be measured and manipulated. This desire for a single measurable feature of gloss evidently transferred to the perceptual domain of study. Despite the fact that numerous papers subsequently identified differences in perceptual experience of gloss, most research concentrated on the standardisation of measurement and the search for a reliable physical index that the visual system could measure or at least estimate.

1.5.3 Additional factors vs. inverse optics

Pfund did, however, acknowledge that there were additional factors involved in perceptual gloss, as it was already established that when observing two materials with identical surface characteristics (and thus ratio of specular to diffuse reflectance), the darker surface would appear glossier. A role for contrast between specular reflection and diffuse reflectance of the surrounding was already evident - yet this was not taken into account in the search for an adequate measurement of physical as against perceptual gloss. It was not until an article published by Hunter (1937) that notions of additional perceptual gloss factors were expanded. This influential paper proposed a number of different aspects of perceptual gloss - and interestingly, did not focus on how gloss was to be measured objectively, but on determining the qualities that should be measured. Hunter outlined six types of perceptual gloss (see Figure 1b-h):

1. Specular gloss - this is defined as the perceived shininess, or the perceived brilliance of highlights. It is the most commonly measured parameter in experiments as an approximation for the physical measurement of perceptual gloss.

2. Sheen at grazing angles - this is the perceived gloss at grazing angles of otherwise matte surfaces (for instance, very smooth, good quality matte paper can have a slight sheen when viewed at low grazing angles).

3. Contrast gloss - identified by contrasts between specularities and the rest of a surface, this is associated with the observed contrast between specular highlights and otherwise diffusely reflecting surface areas.

4. Haze - this is the presence of a hazy or milky appearance, adjacent to reflected highlights. An example of this might be the haze surrounding a reflected highlight on a brushed metal surface.

5. Distinctness-of-reflected-image gloss - this is the perceived distinctness and sharpness of a pseudoimage seen reflected in a surface.

6. Absence-of-surface-texture gloss - this is the perceived smoothness of a surface, where non-uniformities of surface texture such as blemishes are not visible.

Images illustrating these types of gloss can be found in Figure 1. Hunter stipulated that the measurement of gloss should involve one or more of these types, to take into account the additional perceptual differences. He considered the perception of gloss in human vision to be a gestalt (corresponding to no single physical property of a surface, but formed by an appraisal of the whole scene); and that if there were indeed several types of gloss, no one device alone could measure it. In fact, two instruments commonly used to measure gloss in industrial or experimental settings were developed with the intention of measuring gloss in different ways - the glarimeter, or glossmeter, measures the ratio of specular to diffuse reflection, and the Dori-gon measures the distinctness of image - which correlate with two of Hunter's dimensions. By Hunter's description, gloss is more complex than Pfund originally proposed, but is still in some way measurable in objective physical terms.

Despite this, theories proposing a single objective measure persisted; perhaps influenced by pervasive hypotheses concerning the computations involved in human vision generally. The inherent problem in the study of vision is that the information available to the brain from perceptual input is insufficient to provide an adequate account of the surrounding environment - a full representation has to be constructed from the information available. The theory of inverse optics proposes that the brain essentially inverts the sequence of physical processes to reach a model of the environment. Applying this theory to the field of colour vision - the brain tries, according to inverse optics, to calculate the original surface reflectance functions by discounting the illuminant, using reverse physics to approximate intrinsic physical properties of the surroundings. However, this kind of computation would be highly complex and - critically - could hardly ever yield sufficient information to arrive at a solution. A computational model of inverse optics could, however, demand that the brain estimates a single physical objective measure of a property such as gloss, thus explaining the desire to encompass gloss with a single variable which corresponds and agrees with human perceptual judgements. One should not gain the impression that theories based in inverse optics have been completely discarded. In the 1990s Blake and Bülthoff concluded that the visual system 'seems to employ a physical model of the interaction of light with curved

Figure 1.2: (a) shows an advertisement for an Ingersoll Glarimeter, 1922. Reproduced under the Creative Commons license. (b-h) illustrate examples of Hunter's six cues to gloss. (b) shows sheen at grazing angles, on a piece of high quality matte paper. (c and d) demonstrate both surface texture and distinctness-of-image gloss: (c) is focused on the fingerprint-blemished surface, whereas (d) is focused on the reflected image - the surface appears less glossy in (c) as the surface texture of the blemishes detracts from the surface gloss, and the distinctness of the reflected image is lower. (e) shows the original photograph of a shiny surface with a strong highlight. In (f) all highlights have been removed, and the surface looks matte. In (g) the highlight has been reduced to demonstrate contrast shine, and in (h) all haze surrounding the highlight has been removed from the original image.

surfaces, a model based firmly on ray optics and differential geometry' (Blake & Bülthoff, 1990, p165). Their conclusions that the use of specular reflections and their geometry provide rich information concerning the three-dimensional structure of the object are still invaluable even when considered in alternative heuristics frameworks to inverse optics. Inverse optics retains attraction as a basis for theory, despite its intractability. Although clear differences between physical and perceptual conceptions of gloss were evident early in the study of gloss, these were not wholly acknowledged in the search for a perceptual measure of gloss that could be employed by the visual system to identify glossy surfaces and to compare relative gloss.

1.5.4 Emerging support for multiple factors

A gestalt concept of gloss was supported by the work of Harrison and Poulter (1951). This gestalt, they proposed, would include a combination of mainly specular reflection with contrast of specular and diffuse reflection, besides a number of other factors. Later papers developed this, coming from a wide range of research backgrounds. For example, snow was found to have a high contribution of specular reflection at higher angles of incidence, and yet at such angles does not appear shiny - at most, one sees a very bright glare reflected from the snow (Middleton & Mungall, 1952). This is because, considered as material, or 'stuff', the surface of fresh snow is made up of millions of uniquely shaped snowflakes, and the facets of these three-dimensional structures scatter light in all directions (some light is also transmitted through the layers of snow, and partially absorbed). It might be inferred from these results that the microstructure of the surface of the material is also important: the reflection of purely specular light alone does not produce perceptual glossiness. It seems we need a continuous area of the surface to be visible in order to assess the presence of gloss (e.g. smooth sheets of ice look very shiny). An informal paper from the Artificial Intelligence Laboratory of MIT concludes that the perception of glossiness arises as a result of at least two visual effects - that specular reflections from a surface producing mirror-like images of the surrounding environment lie in a different plane from the surface, and that highlights are 'abnormally bright' (Lavin, 1973). Beck and Prazdny (1981) studied such specular highlights more formally, and found that not only are they important for the perception of gloss, but also the orientation and positioning of any highlights are crucial. The size, shape and position of the highlights should be consistent with the three dimensional structure of the object or material, and the supposed angle of illumination. However, the authors also conclude that specular highlights appear to have a purely local effect, that makes only the surrounding area of the surface or object appear glossy.

One of the first attempts to link perceptual gloss with physical parameters of materials was made by O'Donnell and Billmeyer (1986). This paper was a direct consequence of the work of Hunter; reiterating that visual observations led to the identification of six types of perceptual gloss (specular, sheen, contrast, haze, image distinctness, and surface texture). The interrelations between these six types were studied using a multidimensional scaling method, the results of which produced unidimensional interval scales of gloss. However, these scales only appear to apply to the very specific stimuli used, and the particular viewing and illumination conditions. An effect of extreme viewing angle on perceived gloss was acknowledged, but not incorporated into the multidimensional scaling analysis. The physical parameters of the stimuli were analysed using a conventional glossmeter (designed to measure specular contrast) and a Dori-gon instrument (designed to measure distinctness-of-image gloss); other types of gloss - explicitly discussed in the aims of the paper - were not fully considered. (It is worth pointing out here that whilst experiments prior to this refer to the instruments used as 'glarimeters', these are the same as glossmeters, and measure the ratio of specular to diffuse reflection). Two sets of equations for perceptual gloss were produced: each mapped the analysed perceptual responses to the measurements obtained from only one of these instruments. If a glossmeter is used to capture the specularity contrast of the surface (one of Hunter's six dimensions of gloss), which was not found to be independent of lightness, and a Dori-gon instrument is used to capture distinctness-of-image (another of Hunter's dimensions) where the measurements are found to be lightness independent, then Hunter is clearly justified in arguing that specularity and distinctness-of-image are two separate dimensions, and that gloss is not unidimensional. For the glossmeter (but not for the Dori-gon), three linear equations were required to explain all the data (where each equation mapped the unidimensional solution for perceptual responses to a scaled instrument reading), depending on the lightness level of the stimuli. This suggests that unidimensionality is an unusual conclusion at which to arrive - lightness clearly affected the perception of one kind of gloss (as evidenced by the contrast effect, Pfund 1930). Since the stimulus set, viewing and illumination conditions were highly specific, this suggests that even disregarding the problem of lightness, the equations would not generalise to alternative conditions - or even to natural scenes of broadband illumination. This is a particularly important aspect of the study of material properties - rather than searching for computations only useable under specific conditions, the solution needs to be applicable under a wide range of circumstances.

A separate paper by the same authors (Billmeyer and O'Donnell 1987) used magnitude scaling to estimate perceptual differences between all possible pairs of stimuli (using the stimulus set from O'Donnell and Billmeyer 1986). This again produced unidimensional interval scales of perceptual gloss despite apparent consideration of all six dimensions proposed by Hunter. Data obtained correlated with instrumental gloss measurements made with standard glossmeters: but as glossmeters provide a simple ratio of two measures (specular and diffuse light), disregarding a great deal of information, this result is implausible. It seems that the range of information available in the set of stimuli was limited, and thus perceptual judgements of gloss were restricted to the use of specular information, forcing the decisions to be consistent with glossmeter predictions. This provides further support for the conclusion that methods of stimuli presentation and conditions of illumination and viewing were too specific. Bartleson highlighted the need to recognise the multidimensional nature of perceptual gloss in a report to CIE many years previously; yet this was largely overlooked in subsequent work (Bartleson 1974, as cited in Sève 1993).

1.5.5 Persistent support for a single-measure approach

Further studies at around the same time persisted in the assumption that the measurement of gloss - as relating to perceptual experience - could be achieved using a single physical measure. Keane (1989) described in a patent paper the invention of an optical instrument, which could assess both the chromaticity of a surface, by measuring the wavelength reflectance function, and also gloss; the assumption being that colour perception is influenced by perceived surface gloss (U.S. Patent No. 4,886,355). Again, perceived gloss was considered to consist entirely of specular reflection. Considering that the invention was designed to provide a measure capable of compensating for additional factors in perceived colour, it is paradoxical that it neglects evidence in favour of the involvement of multiple factors in perceived gloss. Serikawa and Shimomura (1993), from the field of computer science, went as far as denying the idea that the specular reflection of images of the environment appearing on a different plane from the material surface corresponds with perceptual glossiness. Instead, they defined their measurements of perceptual gloss as involving a brightness function and the smoothness of an object's surface. It is a moot point whether the insistence of industrial research on a unidimensional approach to physical and perceptual gloss may have influenced research in the field of vision more generally. However, their conclusions regarding the measurement of perceptual gloss are in clear agreement - that a single objective scale is sufficient.

1.5.6 A return to multidimensionality

The tendency to cling on to a single-measure approach to perceptual gloss, in spite of the work by Hunter, was finally challenged in a critical review paper by Sève (1993). Many of the problems facing the study of gloss were addressed directly, and attention was drawn to a number of aspects previously neglected. Complications regarding the concept of gloss itself, by this point, were clearly evident. Although Schanda (1971, as cited in Sève 1993) had outlined difficulties with defining and measuring gloss in a memorandum to CIE two decades earlier, this was evidently overlooked by most studies. Even the vocabulary of the CIE definition of gloss shifted from physical to perceptual, without noting explicitly the significance of this change (as cited in Sève 1993). Terms for perceptual and physical concepts were being used interchangeably, so problems of terminology affecting the discussion were inevitable. In the field of colour vision, by contrast, a careful distinction is made between physical and perceptual terms or concepts, preventing such confusion (wavelength, luminance and purity characterise the physical dimensions of colour, whereas hue, brightness and saturation describe the perceptual qualities). In the interest of clarity, Sève adopted the term 'photometric gloss' for visual or perceptual gloss (Sève's term and the later-used 'psychometric gloss' are broadly equivalent).

An important point emphasised by Sève is that the choice of any physical gloss scale is arbitrary, as most instruments make some calculation of specular gloss alone. Yet it is not fully clear how these physical features will best correlate with judgements of perceptual gloss. Sève reiterates the importance of Hunter's multiple visual criteria for determining perceived gloss, and acknowledges that specular reflectance alone does not give a full explanation of perceptual gloss. Appraisal of gloss by the visual system is not dependent on one physical quantity, and does not try to measure or estimate a single physical quantity of the surface reflectance. This is the final nail in the coffin of a single estimated value of the physical world employed by the visual system to approximate gloss; and the theory of combined perceptual factors determining perceived gloss is reinforced.

One crucial point noted by Sève was that visual evaluation of gloss differs considerably from one observer to another. One observer attaches significance to certain characteristics of a scene that another does not, and so samples cannot be ordered linearly. From this fact alone, multidimensionality of perceptual gloss is intuitively inferred, with numerous contributions from different factors. Vision typically involves disentangling information obtained from the environment in the early stages of processing at the retina. For example, effective colour constancy requires the separation of illuminant and surface reflectance, which is further complicated by physiological limitations at the initial input stages of the visual system. All conundrums of vision involve a complex interplay between illumination, object or surface reflectance, and observer. Gloss as a percept is no different; observer, illumination conditions, lightness, contrast, specular reflectance, surface texture, highlights and their properties, specularly-reflected mirror images and binocularity all play a role in the perception of gloss.

Subsequent to this influential paper by Sève, published research on gloss appears to decline for several years. Then in the late 1990s and early 2000s, publications investigating gloss reappear. One such paper seems to signal a change of research tactic - moving from the study of objects, to the perception of materials and surface properties. Adelson (2001) points out that relatively little attention had been paid to the recognition of materials, as opposed to objects - the 'stuff' that makes up what we see is essential for judgements concerning the nature of the object; such as what it might feel like, or how it might be used. This emphasis on the study of textures and material appearance seemed to reignite the study of gloss as a surface property, and encouraged a change in approaches by sparking a variety of new methods (heavily influenced by developments in technology). More recently, studies on the representation of material properties such as texture and colour have also drawn attention to the study of gloss (Cavina-Pratesi et al. 2010b, Cavina-Pratesi et al. 2010a, and Fleming, Dror, and Adelson 2003). In particular, these raise the question of whether the processing of gloss might be independent from the processing of texture, and other surface properties.

The clear assumption from Adelson's paper onwards is that gloss is a complex interaction of illumination, surface, environment and observer. This assumption gave rise to a new range of methods and approaches that take into account the multidimensionality of gloss. Since all problems of vision involve the interaction of illumination, surface, and observer, the findings are grouped accordingly - moving from illumination through surface to observer - including interactions between stages as appropriate. The main aims of the research - describing perceptually distinct dimensions of gloss, computation of perceived gloss from images, evaluation of gloss constancy, and the search for the specific cortical regions involved in gloss perception - are evident throughout the body of findings, and will be flagged as such.

1.5.7 Illumination

Real-world illumination

The importance of realistic illumination distributions in achieving a good level of perceptual constancy is evident in the study of colour vision, and it seems to play an equally important role in surface texture perception constancy. Natural illumination maps have characteristic non-Gaussian fluctuating statistical properties; and so Hartung and Kersten (2002) measured a number of natural illumination maps to investigate potential sources of information for perceiving objects as shiny. Consistencies between illumination of the background environment and the patterns of light reflected from objects were statistically correlated, and any step towards a non-natural map of illumination was immediately salient (a shiny object in an illumination map of white noise appeared matte). This indicates that the complexity of natural illumination maps is crucial for accurate and ecologically valid perception of surfaces, as the visual system takes advantage of this complexity - either through the explicit information available, or by means of correlations and similarities between the surroundings and objects within them.

These results corroborate the findings of Fleming et al. (2003), obtained in matching experiments. An asymmetric matching task was used to measure perceived glossiness for spheres simulated in some comparisons under real world light fields, and in others with geometrically simple illuminants. Observers' matches were close to veridical under geometrically complex real world illuminations, but not under non-natural illuminations (that were not geometrically complex). This implies that complexity of illumination is necessary in the initial stages of surface material perception, and that compensation for a lack of this complexity is not possible later - although constancy is still not perfect under complex illumination. Dror, Willsky, and Adelson (2004) also provide support for the idea that the visual system takes advantage of characteristics of natural illumination maps, arguing that real world illumination is highly complex, and yet possesses a high degree of statistical regularity. If such statistical regularities could be assumed and utilised, this would marginally lessen the complex task of determining the properties of objects in the environment, and would also in part explain failures in perceptual constancy.

Olkkonen and Brainard (2010) found that changing the light field had a significant effect on perceived glossiness, as assessed with a matching paradigm, and concluded that the complexity of computing an estimate of glossiness is increased by a change of illumination. Two physical parameters determining surface properties, diffuse and specular reflectance, were manipulated by the observer across scenes illuminated with different light fields. Gloss constancy was not found across changes in illumination field. Importantly however, the effects of illumination changes on lightness and gloss were different and independent. The current consensus is that many pseudocues[1] are involved in the perception of gloss. Variation in the stimuli of multiple physical cues may well provide more than one pseudocue for glossiness; particularly considering that the stimuli were presented on a high-dynamic-range display, which provides more natural and physically accurate representations (and thus more accurate physical cues for gloss) than the more commonly used CRT screen. However, as the observer only manipulated two physical cues, analysis of the responses is based solely on adjustments of two variables. Thus, while this is an improvement over experiments allowing manipulation of just one variable, and while it may well be the case that a change of scene has a significant effect on perceived glossiness, it must be noted that the results of this particular paper quantified and calculated this effect using two single observer-manipulated cues.

In the same year, Doerschner, Boyaci, and Maloney (2010) took a different approach to the same problem. Pairs of surfaces were compared for glossiness under two different

[1]For the purposes of this review, 'pseudocues' or 'perceptual cues' will refer to cues that the visual system extracts from the scene, and 'cues' will refer to physical properties such as specular reflectance.

real world light fields, and the data used to estimate transfer functions capturing the way in which perceived gloss was remapped from one field to the next. These remappings were best described linearly, and also exhibited transitivity. Some deviations from gloss constancy were shown; however it was found that the nonlinear scale of perceived gloss for one light field was the linear transformation of the nonlinear scale of perceived gloss for another light field. This is a significant discovery, as in many areas of perception the task to be accomplished is often approximated mathematically - which is an efficient and useful tool - however it is not assumed that the visual system might actually be employing a similar technique.

There is some reason to believe that the visual system is not capable of performing such calculations with the information available; yet other findings indicate that such tasks are somehow achievable. For example, in colour constancy, changes in illumination are computationally problematic, as a change in the illumination of a single surface alters the signals given by the L, M, and S-cones. In theory the proportional combinations of the signals given by these cones could differ wildly from those of the initial illuminant, as the proportions of the illuminant light at each wavelength might well be skewed in the opposite direction. However, Foster and Nascimento (1994) estimated L, M and S-cone values based on an illuminant change between two natural illumination maps (skylight and sunlight), and found that the change in L-, M-, and S-cone values could be explained well by multiplicative scaling of the signals, where the relative scaling value differs for each cone class and these values depend on the particular illuminant transition. On a similar note, the conclusion of the Doerschner et al. paper found that a linear transformation can be made between perceptual parameters that are themselves nonlinear (the nonlinear relationship between the physical dimensions of gloss and the perception of gloss). Although such a relationship can be intuitively understood, there is no reason that this should be the case; for this reason it is an important finding.

Motoyoshi and Matoba (2012) carried out further studies of this nonlinear relationship between physical measurements and perceptual judgements of gloss, and found that varying the statistical characteristics of the illumination had systematic effects on perceived glossiness. Thus, while the relationship may not be linear, it is consistent to some extent. (The authors also concluded that judgements of gloss could be predicted by sub-band histograms of the images showing low level image properties - this was disputed, and will be discussed later in this review).

In a more general study of material perception and the effect of illumination (rather than of gloss specifically), Pont and te Pas (2006) found that material perception and light-field perception were essentially confounded in rendered images. However, when presented at a symposium (te Pas & Pont, 2005), these images were recreated with real-world stimuli, adding complex natural illumination. Subsequent judgements of materials were disambiguated, but less so for judgements of illumination. The addition of three-dimensional texture was most helpful in aiding material perception judgements; but this is a useful illustration of the importance of using complex real-world illuminants in obtaining veridical perceptual judgements.

Direction of illumination

The composition of illumination is not the only important component - it is also evident that its direction can have a significant effect on the perception of gloss and texture. Using the relatively new method of Maximum Likelihood Conjoint Measurement, Ho, Landy, and Maloney (2006) varied the illumination direction for surfaces of varying bumpiness. All participants perceived surfaces to be significantly bumpier with decreasing illuminant angle. This was not a failure of discrimination, and additional contextual cues to lighting direction did not improve roughness constancy. Thus it appeared that observers may be relying on features contained in the texture itself (such as highlights, shading and cast shadows) which change with the illumination. This was supported by a study by Nefs, Koenderink, and Kappers (2006), where differences in perceived surface relief were found to result from changes in illumination direction, but not from differing surface properties (glossy or matte). No evidence was found for glossiness influencing shape perception, however - so it seems to be the case that lighting direction influences the perception of texture and surface relief, and not vice versa. Leloup, Pointer, Dutré, and Hanselaer (2010) also investigated whether the geometry of illumination - or luminance contrast - affected gloss perception, and although visual judgements of gloss did not correlate with instrumentally measured specular gloss (as might be expected, from previous discussion), psychometric gloss was a better correlate. However, illumination geometry was again found to be an important factor.

The importance of real-world illumination makes an appearance here, too - Pont and te Pas demonstrated that illumination complexity can manipulate judgements of lighting direction, as well as judgements of surface reflectance (2006). Using a discrimination paradigm, observers' abilities to discriminate between changes of illumination direction and changes in object surface reflectance were explored. This was first performed with computer rendered stimuli, and then with photographs of real objects. Discrimination was not supported with the rendered stimuli, while above chance performance was possible with photographed real-world objects. So again with certain types of rendered image, some cues important for perceptual judgements are evidently being omitted - the most salient being real-world illumination distribution.

1.5.8 Illumination and object/surface interaction

Specular reflectance

Specular reflectance does not consist of specular highlights alone; but all light reflected from a surface where the angle of incidence of the light and the angle of reflection are equal. This is one of the many cues that have been proposed as potentially informative in the perception of gloss, as glossy objects have a higher proportion of specular to diffuse reflection. There is support for this argument, as subjects can judge the specular reflectance of computer simulated glossy surfaces (Nishida & Shinya, 1998), and can also estimate particular properties of the surface reflectance without access to explicit information about the illuminant (Dror, Adelson, & Willsky, 2001). The solution for this, proposed by the authors, is that we rely on statistical regularities in the spatial structure of real-world illumination; and that these regularities are sufficiently predictable to allow us to estimate surface properties from statistical features of the image. This is consistent with both the gestalt view of perception, as well as the 'bag-of-tricks' computational approach (Ramachandran, 1985).

Specular highlights and their properties

As a result of numerous studies, it is now recognised that a number of properties of specular highlights must be present for gloss to be perceived convincingly. These properties include the relative brightness of the highlights, their contrast, position, orientation, and consistency relative to the object surface and shading.

An early paper on the properties of highlights found that increasing their size and brightness increases the area of the surface perceived to be shiny (Beck & Prazdny, 1981). The orientations of the highlights are also important - they must lie in the direction of minimal curvature, and the perceived gloss increases if they are consistent with the intensity gradient of the surface or of the surface contours (that is, the three dimensional shape information). This was supported by Hurlbert, Cumming, and Parker (1991), with the finding that increasing the brightness of the highlights increases the perceived level of gloss. Marlow and Anderson (2013) also showed that objects appear glossier if images are generated with a higher specular coverage; with increased sharpness and contrast.

Not only must highlights have certain properties in terms of relative brightness, sharpness and contrast, but they must also be consistent with the three dimensional shape of the object overall. For instance, the shading of an object should be congruous with the three dimensional shape in terms of the lines of contour; changes in illumination help to resolve any ambiguities in the solid shape of the object (Koenderink and van Doorn, 1980). Highlights placed on a two-dimensional image of a vase with shading consistent with the supposed three-dimensional geometry give a good impression of surface gloss if compatible with the lines of contour (Beck, 1964). Specular reflections also provide reliable and accurate constraints on the three dimensional shape, as there is a distinctive and characteristic way in which the reflected light (and the pseudoimage) is warped across the surface of the object, compatible with the three dimensional shape (Fleming, Torralba, & Adelson, 2004). These specularities can be distinguished from differences in texture, and remain consistent even with changes in environment. They can be extracted by populations of simple oriented filters. However, even when these conditions are met, the gloss ratings given by observers are not uniform across a surface with highlights (Berzhanskaya, Swaminathan, Beck, & Mingolla, 2005). Gloss ratings decrease as a function of the distance from a highlight, even when the distance is discounted from luminance values. This finding suggests that gloss constancy is restricted to a local level. The visual system does not appear to operate under the assumption that glossiness remains constant over a single object, unless there are similar reflections across the entire surface - which might be a possible flaw of gloss constancy. However, this would explain why objects rendered under realistic illuminations rather than single-point light sources look more glossy (Fleming et al., 2003), as the illumination geometry is more complex, giving a broader spatial distribution of light on the surface of the object. This means that more highlights, or local gloss percepts, are generated across the surface of the entire object.

Kim, Marlow, and Anderson (2011; 2012) supported the notion that multiple facets of specular highlights need to be considered. Besides concurring that highlights should be congruent with surface shading, they further suggested that the perception of gloss does not depend on the brightness of highlights alone; but that the locations of the specular reflections must correspond to the diffuse shading profile of the surface. This was demonstrated by adding lowlights - rather than highlights - to matte images, which gave as convincing a perceptual experience of a glossy surface as adding highlights. So it seems adding either highlights or lowlights can give the impression of gloss - combined with sharpness and contrast of highlights. Marlow, Kim and Anderson (2011) also investigated the relationship between highlights and the diffuse shading profile, and varied highlight orientation relative to the diffuse shading of the surface by rotating the highlights. The distance of the highlights from the brightest region of diffuse shading was also varied, by transposing highlights in displays while also preserving the orientations of the highlights relative to their surrounds. Previously, highlight incongruence had been generated by simultaneously displacing the position and orientation of highlights in the image. It was therefore important to try and separate these two variables, to ascertain whether only one or both variables affected the judgments. Manipulating either variable in a non-natural direction reduced the perceived gloss; although rotations reduced perceived gloss more than transposed highlights, despite the fact that this displaced highlights into darker regions. Together, these findings provide further support for the view that the perception of surface gloss depends on highlight congruence with the structure of diffuse luminance variation in an image, and not just consistency with surface shape. While the highlights must be congruous with the diffuse shading profile of an object, the highlights themselves do not appear to influence perception of the diffuse shading profile. Todd, Norman, and Mingolla (2004) found that observers can discount the presence of specular highlights so that the relative lightness among different regions of the image is determined almost entirely by the diffuse component of surface reflectance.

However, while highlights do not seem to affect perception of the diffuse shading profile, the presence of specular highlights does bias judgements of ambiguously shaded objects towards a convex interpretation (Adams & Elder, 2014). This effect is likely to be an assumption based on illumination geometry, as highlights are less likely to appear on concave than convex surfaces. The effect decreases if the highlights are misaligned with regard to the surface shading, as they are more likely to be perceived as a feature of the surface rather than as a specular highlight.

Interactions of object surface and illumination play a significant role in the perception of gloss. Marlow, Kim, and Anderson (2012) proposed that changes in perceived gloss could be understood as a direct consequence of image properties that covary with surface geometry and illumination field. A change in either of these factors can generate different patterns of interaction with perceived gloss, and these interactions can be complex and variable. However, Marlow et al. argued that the successes and failures in the perception of gloss can be predicted by the way that each illumination field modulates the characteristics of the specular reflections. Such effects provide strong evidence for the modulation of perceived gloss occurring as a direct consequence of a systematic covariation of specular reflections with changes in the distal scene. However, to judge perceived gloss in this study, Marlow et al. used the variable with the largest apparent difference between stimuli - either the degree of coverage of specular reflections, sharpness, or contrast. This might suggest that the visual system makes a judgement based on a number of different types of information, where each does not contribute in a consistent way to the overall experience of gloss. This could provide an explanation for the supposed instabilities in perceived gloss when changes occur in surface geometry or viewing conditions. If the relations between physical parameters and perceived experience are nonlinear, perceptual features may vary in salience depending on the manipulations made, such that judgements would be made on the basis of different perceived variables each time.

1.5.9 Object properties

Surface texture and shape

The three dimensional shape of an object can affect perceived gloss alone, as well as through interaction with changes in the field illumination. Marlow et al. (2012) showed that perceived gloss of a surface varied up to 80% as a function of the three dimensional surface relief alone, within a single illumination field.

Furthermore, there appears to be a significant influence of shape on the perception of material reflectance. Vangorp, Laurijssen, and Dutré (2007) found that when comparing two objects of identical material where the geometry of the two differs, accuracy of material perception decreases. The addition of edges significantly changes the perceptual judgement of the material; and two different materials presented in the same shape can look identical despite having very different reflectance properties. For example, two tessellated spheres rendered with two different types of blue plastic appear to be identical, and two objects rendered with identical materials but in different shapes (a smooth blob, and a tessellated sphere) are perceived very differently. The blob-shape appears to be very glossy, mainly as a result of the curved surface displaying a range of highlights, while the tessellated sphere mostly reflects diffusely and is perceived to be made of a matte material. (All images were rendered with real-world light probes, so specificity of a limited or unnatural illuminant did not affect judgements). This finding is supported by a study by Nishida, Motoyoshi, and Maruya (2011), where observers were found to have limited ability to recover surface reflectance properties under changes in surface shape - indicating that three-dimensional object shape can influence our perception of surface gloss. Olkkonen and Brainard (2010) found that both shape and illumination affected perceived glossiness, and that there were large interactions between illumination and object shape in their effects on perceived glossiness. Joint effects of the individual factors could not be predicted from the individual effects in a straightforward manner, and analysis of luminance histogram statistics could not account for the interactions. This can be related to the findings of Ho, Landy, and Maloney (2008) in terms of the use of ‘pseudocues’ - both shape and illumination field affect the pseudocues, yet the translation from physical measurements to pseudocues is not necessarily linear or even monotonic. The mechanisms may interact with each other in a nonlinear way in physical terms, or the perceptual pseudocues translate from the physical in a nonlinear manner. To date, these effects remain unexplained.

Surface properties other than the shape of the object itself play a further role in the perception of gloss. Ho et al. (2008) demonstrated that variation in three dimensional surface texture significantly affects gloss constancy: if a surface texture is bumpier, this results in an increase in perceived gloss. However, beyond a certain level of bumpiness (with a large difference between the high peaks and low troughs) the surface looks less glossy. This study was performed using a conjoint measurement paradigm, and Ho et al suggested that the observed interactions between perception of gloss and bumpiness of surface texture are the results of imperfect cue learning (or use of pseudocues - that is, indirect use of the physical information available).

These conclusions were partially supported by Qi, Chantler, Siebert, and Dong (2012) who studied how mesoscale and microscale roughness affect perceived roughness. Mesoscale roughness is of a lower spatial frequency than microscale roughness - that is, the 'bumps' themselves are of larger size. (As an example, mesoscale is to microscale as pebbledash is to sandpaper). Perceived gloss changed monotonically when varying the microscale roughness parameter, and non-monotonically when varying the mesoscale roughness parameter: that is, both parameters affected perceived gloss, yet an additive model was inadequate to describe the interactive and nonlinear influence. As in the study by Ho et al. (2008), the effect of surface texture was non-linear, and changes in approximately mesoscale roughness did not produce a consistent effect on the perceived glossiness.

Surface lightness

Surface lightness, regardless of colour, also has a significant effect on perceived gloss - and there seems to be an effect in both directions. Harrison and Poulter (1951) observed that dark surfaces appear glossier in comparison to lighter surfaces. Glossier surfaces appear darker than their rough/matte counterparts, apparently due to increased contrast between the specular and diffuse components (Beck, 1964). However, this seems to conflict with Todd's (2004) findings that observers are able to discount the presence of specular highlights in determining relative lightness. Todd et al. concluded that observers were able to exclude specular highlights in making their judgements of relative lightness, which were subsequently determined largely by diffuse reflectance from the surface - although this apparent conflict is based on the assumption that glossiness is entirely determined by specular highlights.

Surface colour

Research to date has produced conflicting and uncertain results regarding the potential influence of surface colour on gloss perception. Initial studies seemed to show that there was little in the way of an interaction - Xiao and Brainard (2008) found little evidence to suggest that variation in surface gloss had a noticeable effect on the appearance of colour. In one condition, surface gloss and body colour of a sphere were varied, and in the second condition the point on the object at which the participant observed the colour was varied. The visual system seemed to compensate for the physical effect of varying gloss, but a small effect was still observed on the perceived colour appearance. An effect of patch location was also found, though smaller than the physical effect, but compensation of test patch location did also occur. However, later studies found some evidence in favour of colour information affecting gloss perception. Wendt, Faul, Ekroll, and Mausfeld (2010) showed that the inclusion of colour information in stimuli improved gloss constancy performance (although gloss was classified using only specular highlights). Availability of colour information led to a significant improvement in consistency in glossiness matching (that is, fewer systematic errors) compared to greyscale surface trials. Some observers even gave priority to colour information over motion (discussed in the next section) as a cue to glossiness; although in general observers showed different levels of receptiveness to certain combinations of information to be used in making a judgement. This implies at least some basic input from colour processing, but again indicates that different observers prioritise different cues for gloss.

More recent investigation into the potential importance of colour processing for perceived gloss has focused on colour information obtained from the specular and diffuse components (Nishida et al., 2008). When wavelength compositions of specular highlights and diffuse light were changed, observers perceived naturalistic glossy surfaces only when the physical constraint of the highlight held. In other words, highlights comprise a wide range of wavelengths of light, including the surface reflectance and the illuminant. The diffuse component, however, cannot contain any wavelength absent from the reflected highlights, as this is composed of all wavelengths in the illuminant, and additional wavelengths cannot be added when reflected from a surface. Gloss perception was also reduced when there were no luminance increments between the diffuse reflectance and highlights. A subsequent paper (Nishida et al., 2011) found that multiple colour band analysis using raw cone-signal based images could not fully explain the luminance-colour interaction in gloss perception, as when an image synthesised from S, M, and L cone images violated the physical constraint it was still perceived to be naturally glossy. However, the authors concluded that this kind of multiple colour band analysis might be a promising hypothesis for observed colour and luminance interactions. In a similar study (Hanada, 2012), the colour coordinates of the objects and highlights were varied, while luminance was unchanged. Objects were perceived as glossier when the highlight and object colours were different, demonstrating the normal difference between purely specular reflection and surface reflectance. Unnatural combinations of colours were still perceived to be relatively glossier, when compared to stimuli with identical surface and highlight colours, even though the luminance of each pixel of the images was controlled.

1.5.10 Observer and object/surface interaction

Motion information

When we perceive objects in everyday life, we are not limited to viewing static objects. We are continually moving around our environments; and if not changing our physical position, we are constantly making eye saccades. This motion produces a steady stream of optic flow, which provides a rich source of perceptual information about our surroundings. When inspecting a new or interesting object, we might pick it up and rotate it by a window. Such inspection allows us to investigate the surface properties of the material, by observing the changes in surface reflection. Hurlbert et al. (1991) noted that specular highlights appear to remain stationary on the surface of a rotating sphere when the observer is stationary: - the highlights appear to slide across the surface of the object, and thus remain stationary relative to the observer. It is evident that a great deal of information about surface properties such as gloss can be extracted - the movement of specular reflection across an object reveals a great deal about its three dimensional shape, and this movement is particularly revealing for glossy objects. Hartung and Kersten (2002) showed that the pattern of optic flow projected from a rotating shiny object is significantly different from that of a rotating matte object. A number of objects were 'painted' with the image of an illumination map, so that for any given static view it appeared shiny - but when it began to rotate, it appeared matte. Rather than staying stationary, the specular highlights moved with the surface of the object - thus producing a different pattern of optic flow.

These findings are well supported. Sakano and Ando (2008) investigated the effect of self-motion through a scene on gloss perception. Temporal changes in the scene caused by lateral motion of the observer enhanced the strength of perceived gloss; even though rendered stimuli were used. Stimuli on a screen moved in accordance with any movement of the observer's head, to simulate movement in a three-dimensional space. The stimuli luminance also changed temporally in terms of the spectral highlights as well as position, so that the object appeared to be stationary in 'three-dimensional' space, while a reference stimulus did not change on the monitor. Similarly, Wendt et al. (2010) found that motion information significantly improved gloss constancy performance - systematic errors were significantly smaller in gloss matches under dynamic conditions compared to static conditions, regardless of whether binocular information was available. (This is readily confirmed by real-life situations, when we rotate objects in our hands to see highlights move across the surface - while remaining stationary relative to the illuminant - to assess glossiness). Doerschner et al. (2011) investigated whether there might be a characteristic way in which such features move during object motion or changes of viewpoint, which might act as a reliable source of information in judgements of gloss. For moving stimuli, subjects reported that objects with normal specular motion appeared shinier than those with static reflections (relative to the object). However, on trials where the object did not move, performance was at chance level - indicating that motion cues alone caused differences in appearance, rather than the way in which the motion stimuli had been created. Rather than just contributing to the perception of glossiness of an object, these motion cues could be used to distinguish between matte and shiny surfaces. Therefore the visual system appears to rely on characteristic optic flow patterns in determining glossiness. Lichtenauer, Schuetz, and Zolliker (2013) supported this further in a study where judgements of rough and glossy surfaces were compared, by either interacting or passive observers.

Active exploration of the rendered stimuli gave significantly higher inter-observer agreement of perceptual judgements; supporting the conclusion that the motion of an object, whether facilitated by the observer or the object, reveals a characteristic optic flow which can inform perceptual judgements of gloss.

Ho, Maloney, and Landy (2007) also investigated the effect of viewpoint on perceived gloss, by carrying out an adjusted version of the earlier conjoint measurement study. Bumpiness and illumination were kept constant, and observers were asked to make judgements of the surface properties from two different viewpoints. Observers failed to achieve roughness ('bumpiness') constancy based on similar pseudocues to the previous study, suggesting that the human visual system does not always select the right cues for the visual task. This might seem to contradict the results discussed above. However in this study there was no explicit observation of the transition between viewpoints but rather a comparison of judgements from two locations. It seems to be the case that a change in viewpoint without observing optic flow confuses our roughness constancy, while the inclusion of motion improves it.

Viewing distance

To date, there has been little research into the effect of viewing distance on perception of gloss. However suggestions have been made regarding reasonable viewing distances when conducting empirical studies involving perceptual judgements of gloss. Czepluch (1976, as cited in Leloup, Obein, Pointer, and Hanselaer 2013) recommended that restrictions should be placed on relative distances between the illuminant, object, and observer in gloss scaling in particular, as 'any standard geometry for visual evaluation of gloss [was] lacking'. Such recommendations might be based purely on speculation that increased viewing distance affects perceptual acuity - for instance, Ho et al. (2008) showed the increased bumpiness of a surface alters the perceived gloss. Viewing surfaces of reduced bumpiness, but at closer viewing distances, might mean that observers are better able to perceive a finer scale of texture, which would influence the gloss judgement (Qi et al., 2012). Little is known about this potential factor, but it is undeniably an important variable to control.

Binocular disparity

Even before any detailed study of the perception of gloss began, binocular disparity had already been identified as a potentially invaluable source of information. Kirschmann (1895, as cited in Wendt, Faul, and Mausfeld 2008) proposed that the disparity of highlights on specularly reflecting surfaces usually differs from the disparity produced by points on the surface itself. Czepluch (1984, as cited in Sève 1993) also emphasised the importance of binocular disparity. Highlights reflected from an object appear to be positioned differently to each eye; thus each receives different information about the position of the highlight on the surface, as specular reflection is always reflected - by definition - at an equal but opposite angle to that of the illuminant, and this angle will be slightly different for the two eyes are they are laterally displaced. Thus highlights can be correctly identified, rather than seen as differently coloured patches on the surface.

The importance of binocular disparity has been confirmed by a considerable number of more recent studies. Hurlbert et al. (1991) found that binocular disparity of specular reflections can override brightness in judgements of gloss, and Obein, Knoblauch, and Viénot (2004) proposed that retinal disparity plays an important role in the perception of gloss - mainly in the judgement of high gloss values (i.e. in highlights). While the latter has yet to be investigated further there is substantial evidence for binocular disparity as a significant cue for the perception of gloss. When information from a disparity is available, it can signal that a surface is glossy (Formankiewicz & Mollon, 2009), and perceived gloss appears to be stronger and more authentic (Wendt et al., 2008). In the former study, the author underlined that as the illuminant is directional, not only is there a disparity in the position of the highlight, but the intensity of the reflected light is also slightly different. The angles of reflection to each eye are not identical, yet for light to be specularly reflected the angle of illuminant direction and angle of reflection must be the same. A patch which appears to be reflecting largely specular light to one eye will reflect slightly more diffuse light to the other. Thus, the visual system is exposed to discrepancies in the monocular luminance of highlights as well as their relative location when viewing a glossy surface. The ability of subjects to detect a binocular luminance disparity was measured, and the results were consistent with Weber's law[2], Ricco's Law[3], and Bloch's Law[4], demonstrating that the visual system is more than capable of distinguishing these disparities.

Furthermore, Wendt et al. (2010) showed that the presence of disparity information significantly improved gloss constancy performance, both alone and in conjunction with additional information such as colour and motion. However, developing Formankiewicz and Mollons' earlier findings, Methven and Chantler (2012) found that while stereo disparity increased the perceived glossiness for rough surfaces, specular highlight disparity alone was not enough to ensure increased perceived glossiness. More naturalistic renders of objects and surfaces were used, and the conclusions further confirm the emerging picture of the need for a number of interacting factors. Naturalistic specular highlights are generally sufficient for gloss perception, as long as they are placed correctly, but the constancy of this perception is strengthened by the addition of information such as a disparity in specular highlights. While a single type of information may induce some level of perceived gloss, alone it does not ensure maximal perception of gloss. Evidence supporting binocular disparity by means of a performance based task was obtained by Muryy, Welchman, Blake, and Fleming (2013). Images of specular objects were binocularly presented, and observers were asked to adjust the positions of a number of 'probe dots' to indicate the level of depth that they perceived in the image. For simple surfaces, where there was no indication that the disparities presented were 'wrong', participants erroneously said that the virtual surface was real, by indicating a more realistic level of perceived depth. However, when surfaces were more complex, participants made fewer errors, and correctly identified surfaces with larger disparities in unexpected locations, by indicating much lower values of perceived depth. This suggests that the visual system assesses sensory signals for relevance and usefulness, based on intrinsic markers of reliability. These markers are in the disparity signals themselves, as errors were made at face value - this suggests that the brain interprets specular objects by applying a general strategy instead of implementing physical rules of

[2]Weber's Law states that the Just Noticeable Difference (JND) between two stimuli is proportional to a constant ratio of the magnitude of the stimuli.

[3]Ricco's Law states that, for stimuli of less than one arc-minute in diameter (the resolution of the eye at the fovea, larger at the periphery), spatial summation applies - the threshold intensity multiplied by the area equates to a constant for a test patch to be detected (that is, a larger patch of lower luminance is just as detectable as a smaller patch of higher luminance).

[4]Bloch's Law describes temporal summation, and states that within a certain time limit (100 milliseconds), the minimum number of quanta required to detect a test patch is constant, regardless of whether the patch was of high luminance and lower presentation time or low luminance and higher presentation time (that is, light intensity multiplied by time presented equals a constant for detection of a patch).

specular reflections, which proves useful when the disparity signals are abnormal.

An additional study by Kerrigan and Adams (2013) tested observers' abilities to use specular information and binocular disparity to identify the curvature (convex or concave) of an object, to determine whether this might be invoked by knowledge of a geometric model of specular reflection. Binocular vision enables observers to distinguish specular highlights from other variations in luminance, as unlike surface markings, specular highlights 'float' on a plane above the surface if concave and below if convex. However, observers' performances were not consistent with a full geometric model of specular reflection - showing substantial errors particularly for concave surfaces. Kerrigan and Adams came to the same conclusion as Muryy et al.; that the visual system seems to invoke a general strategy, rather than responding based on an understanding of the physics of specular reflections. However, it is important to note that this is not the same as a 'bag-of-tricks' approach but instead halfway between this and a reverse optics/physics approach.

Physical interaction

Besides interaction with objects in the environment on a purely motion-based level, active handling also appears to improve our visual perception. This might not be limited to motion based information alone - for when we pick up an object to inspect it, we also make judgements of texture using our sense of touch. It is intuitive, but not necessary, that these tactile judgements might feed into visual perception. Bergmann Tiest and Kappers (2007) found that judgements of rough and glossy surfaces were slightly better and more consistent when observers also made haptic judgments, compared to judgements made on the basis of visual observation alone. Interestingly, participants ordered the samples according to different criteria - some ordering on high spatial frequencies, whereas others ordered on low spatial frequencies. This provides evidence not only for a holistic account of texture judgements in terms of the senses, but also evidence for a constellation approach for visual cues (pseudocues). Each observer may give different weightings to the types of information available in making these visual judgements - perhaps based on the kinds of surfaces they have previously experienced.

1.5.11 Observer

Linking the perceptual and physical dimensions

Perhaps the hardest task in this field is the problem of bridging the gap between existing knowledge of the physical dimensions, and perceptual judgements of the human observer. We have already seen that the relationship between the physical and perceptual dimensions cannot be described linearly, but that a linear change in the physical dimension can correlate to a linear change within the perceptual dimension (e.g. Doerschner et al. 2010). Despite our lack of knowledge of the relationship between the physical and the perceptual dimensions, there is evidently a great deal of consistency in the way in which the physical environment is interpreted by the visual system. This is evidenced by findings such as those of Doerschner et al., and also our general day-to-day experiences (visual constancy is sufficiently successful to the point that failures are unusual and interesting).

One of the first experiments to address the problem of linking perceptual and physical dimensions was by Ferwerda, Pellacini, and Greenberg (2001). Here, a psychophysically-based light reflection model of surface gloss perception was proposed; and experiments were conducted to explore how physical parameters describing reflectance properties of glossy surfaces might link to the perceptual dimensions of the appearance of gloss. Multidimensional scaling techniques were employed to incorporate the acknowledged multidimensional nature of gloss perception. As a result, Ferwerda et al suggested that there were two 'perceptually meaningful' axes of perceptual gloss-space: the apparent contrast of a reflected image, and the apparent sharpness or distinctness of this reflected image. Magnitude estimation was then used to place quantitative scales on the axes proposed. However, some concerns about the method should be raised. Participants were asked to judge the apparent difference in gloss in a pair of stimuli by means of a sliding scale ranging from 0 to 100. Such a measure is less reliable in terms of consistency between participants, or indeed even within the judgements of a single participant. When considering a large number of comparisons between stimuli, any given scale needs an established reference point. Differences in pairs of stimuli should be compared directly with other stimuli; otherwise the judgements made cannot be reliably related to one another. A method involving a comparison of two pairs of stimuli, where all possible comparisons within the stimuli set are used, might be more suitable for such an investigation. Observers are asked to indicate which pair they perceive to have the larger difference in the required variable. This would allow the data to be interpreted and quantified in a valid way; and would thus be far more informative. In addition, the reliance on the two proposed axes alone does not allow for any interaction with factors previously acknowledged as influential in the perception of gloss; curiously limiting the scope of further multidimensionality after employing multidimensional scaling techniques.

More recently, Obein et al. (2004) used a maximum likelihood difference scaling paradigm to estimate gloss scales for a series of black coated stimuli. A nonlinear relation between gloss percept and instrumental specular gloss values was found, and sensitivity was higher at extreme scale values than in the middle. If a reverse optics method were being employed, one would expect to find a linear relationship between the percept and instrumental values, as the physical scales themselves would be estimated. Therefore, this nonlinear relationship supports a conclusion favouring a pseudocue- and interpretative-based approach. However, in line with the previous convention, judgements of gloss were reliant only on specular highlights. This shows a non-linear relationship between the physical and perceptual parameters of a single source of information, influential in the perception of gloss. Of course, these initial experiments necessarily manipulated a limited number of variables as they were the first of their kind. Expanding this to incorporate additional variables which factor into our perception of gloss would be a considerable and extremely complex task, yet it is important to note that the information available to observers in this particular case was constrained.

Gloss constancy

A number of different physical and perceptual cues which influence constancy of perceived gloss have already been discussed. Deviations from gloss constancy are evident under a number of different viewing conditions - strong interactions between object shape and illumination geometry produce failures in gloss constancy (Olkkonen & Brainard, 2011), perception of gloss is not independent of light field (Doerschner et al., 2010), and constancy is affected by viewpoint (Ho et al., 2007) and variation in surface texture (Ho et al., 2008). Constancy improves under natural illumination, although is not perfect (Dror et al., 2004;

Fleming et al., 2003), and also improves with the inclusion of colour, motion and disparity information (Wendt et al., 2010). It also seems that gloss constancy operates at a local level (Berzhanskaya et al., 2005). Considering the body of findings related above, it is evident that there are a number of failures of constancy and inconsistencies between physical measures and perceived gloss that are difficult to explain. If the function of perceived gloss is assumed to be identifying surface properties, then constancy of perception is important. There is some evidence for a less-than-perfect gloss constancy (such as consistency of judgements under different illuminants, Fleming et al. 2003), and findings suggest that it operates in a similar way to colour constancy. When more information is available to the visual system in the scene, and when the stimuli are more realistic and lifelike, observers show a greater degree of constancy (Kraft & Brainard, 1999).

A great deal of research has been done on colour constancy but comparisons of colour- and gloss-constancy are not straightforward. The measurement and quantification of gloss constancy involves problems that do not arise in work on colour constancy. In colour constancy both perfect constancy and perfect inconstancy can be objectively characterised (judgements with perfect inconstancy are determined purely by the spectral composition of light reaching the eyes rather than solely by the colour reflectance properties of surfaces). It is not clear what form of judgment would constitute perfect inconstancy for gloss perception; the lack of a 'low-end' to the gloss constancy scale makes it difficult to quantify and compare deviations in gloss constancy. While it can be clear that observers are not achieving perfect constancy in experiments, quantifying the degree of imperfection of judgements in a particular task or comparing deviations across tasks is not generally possible. Of course the relative degree of constancy found when a single factor is varied in an experiment can still be measured. However, as soon as two factors are varied making comparisons between their effects and quantifying their interaction is problematic. Evidence to date indicates that observers are capable of making relative judgements of gloss when comparing stimuli varied along a single dimension, but when multiple factors are jointly manipulated interactions occur and often result in a confound - although these are reduced when motion, complex illumination and colour information are available. It seems that perceived gloss is related to its physical determinants nonlinearly, or at best imperfectly.

Cortical processing of gloss and other surface properties

While much of the focus in gloss research has been directed at the perceptual cues involved by means of psychophysical experimentation, additional lines of enquiry have looked into the processing of perceptual information beyond the retina and into the cortex. Such investigation aims to discover the way in which perceptual information is processed in later stages of visual perception, in order to test theories of essential computations that might be performed; and whether this involves additional unknown factors or processes. This kind of knowledge might feed back into research at earlier stages of visual processing, by highlighting additional perceptual tasks that might contribute to other visual processes. In 2007, Cant and Goodale carried out an fMRI experiment to investigate the cortical mechanisms underlying the roles of object form and surface properties in object recognition. The results suggested that there were different pathways in extrastriate cortex for the processing of form and surface-properties. It was also concluded that the extraction of surface colour seemed to occur relatively early in visual analysis, compared with the extraction of surface texture. A tentative inference from this might be that the extraction of surface texture requires further (and more complex) computation than colour.

In a more recent set of studies, Cavina-Pratesi et al. replicated these findings (2010b). By studying visual object agnosia patients, a behavioural double dissociation was found with a double dissociation in the damaged areas of cortex - one patient could distinguish object shape but not texture, and a second could distinguish texture but not shape. Separate processing of surface texture and form was found in the ventral stream; surface texture activated an area quite distinct from areas activated by shape and form. This is evidence that these areas play a causally necessary role in the discrimination of these features; and that the two tasks are to a great extent accomplished independently by the visual system. In a second paper, Cavina-Pratesi et al. (2010a) sought to determine whether there was a single region involved in the processing of surface properties, or whether there was a number of more specialised regions implicated; each dealing with a particular surface property. A double dissociation was found between two patients, in the processing of surface properties (texture and colour) and geometric (shape) properties. Separate foci were also found for colour and texture - areas selective for shape, texture, and colour were found to be distinct from areas responding to a combination of these features. Thus, it suggests that there are separate channels for processing form, texture, and colour; as well as the division between surface properties and object shape/form.

Kentridge et al. (2012) developed this line of enquiry further with an investigation into whether glossiness perception was mediated by the same processes as colour or surface texture. Gloss is conceptually distinct from texture and colour, but not necessarily distinct in visual processing - yet it was found that glossiness perception could be mediated independently of cortical processing of colour or texture. Patient MS displays a number of visual abnormalities, and is a cerebral achromatopsic - he is unable to discriminate colour and texture, as a result of a lack of these cortical areas. MS performed significantly better than chance on a gloss perception task, for real and rendered stimuli, though slightly worse than controls. This task could not have been solved on the basis of local feature comparisons, as lightness and texture were both randomised. Thus, it was concluded that the perception of gloss does not depend exclusively on processing in the same constellation of regions necessary for the perception of colour and texture.

1.5.12 Neural selectivity

A number of recent studies have investigated the neural correlates of perception of surfaces and their properties, with a small number focusing on the perception of gloss. Results from previous studies (Cavina-Pratesi et al., 2010a, 2010b) suggest that information concerning surface properties is processed in the ventral visual stream, and the results from the studies on gloss corroborate this.

Nishio, Goda, and Komatsu (2012) were the first to investigate neural selectivity for the perception of gloss. They examined the responses of neurons in the inferior temporal (IT) cortex of macaques while presenting stimuli of objects varying in specular reflection, diffuse reflection, and roughness. Neurons in the superior temporal sulcus selectively responded to specific types of gloss - this remained constant when the shape or illumination of the object was altered and perceived gloss was the same, but changed when the images were scrambled and perceived gloss was different. For instance, one cell responded selectively to stimuli with very sharp highlights, and did not respond at all to weak glossiness. A second responded strongly to shiny surfaces that had blurred highlights, and a third responded only to matte stimuli with very low specular reflectance. Nishio et al concluded that there is a population of cells that represent different types of gloss, each cell having a different selectivity. They also proposed that mechanisms in the visual cortex integrate local features of the image to extract information about surface gloss, and that this information is systematically represented in the population of neurons in the IT cortex.

Shortly afterwards, Okazawa, Goda, and Komatsu (2012a) investigated selective responses to glossiness using fMRI. Specular reflection alone was manipulated in generating images of specularly reflecting and matte objects. A set of scrambled images was also produced, and responses to the specular images were compared with responses to the matte and scrambled images. Activation was found throughout the visual pathway, from V1 to V4, and the posterior inferior temporal cortex (only slightly different to the superior temporal sulcus, as found in Nishio et al.). Contrasts of the images were subsequently manipulated, and the activations observed could not be explained by the use of global or local contrasts. Okazawa et al. concluded that processing of specular images occurs along the ventral visual pathway, to particular regions in the IT cortex. This is consistent with the findings of Nishio et al, and also with previous studies of the processing of surface properties in human fMRI - showing that even though specular reflection of the objects was the only variable manipulated by Okazawa et al., their results generally supported previous findings.

Wada, Sakano, and Ando (2014) performed the first human fMRI study on the areas involved in perception of gloss in the human cortex. Given this was a human study, a particular point of interest was that areas beyond the ventral visual cortex have been implicated in processing gloss. As described earlier, Kentridge et al. (2012) found that patient M.S., a visual agnosic with lesioned ventral visual cortex and intact dorsal visual cortex, was able to distinguish between glossy and matte objects at above chance levels. Furthermore, many visual features have been shown to influence human perception of gloss in psychophysical experiments, so plausibly a number of regions could be involved rather than a single localised area. First, they investigated which cortical regions might be involved more generally, by comparing responses to high and low gloss objects. All regions showed significant correlation with perceived levels of gloss, and were consistent with regions identified in the macaque studies apart from V3A/B in the dorsal visual pathway. It was proposed that the involvement of this region could be specific to the human visual system, supporting the findings of Kentridge et al. In a second experiment, visual areas modulated by selective attention to gloss were investigated. All regions showing activation were among those identified in the first experiment. Wada et al concluded that a number of commonly identified regions of visual cortex may be involved in central processing of glossiness, with additional regions contributing to the processing of gloss cues; of which some may be specific to the human visual system.

1.5.13 Subsequent throwbacks to a single obective measure or approximation employed by the visual system

Despite the emerging consensus for a multidimensional account of the perception of gloss, the conclusions of a number of papers hark back to early research. However, the aims tend to the opposite end of the scale of solutions, as a number of 'bag of tricks' approaches are proposed - what might be seen as shortcuts 'that just work' - though in fact none of these have proved especially successful.

Perhaps the most prominent of these attempts was by Motoyoshi et al. (2007) who proposed that there were simple image statistics which could identify perceptual gloss in real-world surfaces. Images of glossy surfaces were analysed, and Motoyoshi et al. found that the skew of the luminance histogram and the skew of the sub-band filter output were correlated with perceived surface gloss - and inversely correlated with diffuse reflectance and a perceived matte surface (where a positive skew correlated with perceived gloss, and negative skew correlated with a matte surface). This was presented as evidence that human observers might estimate statistics such as the luminance histogram skew; in conjunction with evidence that a visual aftereffect was found based on this skewness. Adaptation to images with skewed statistics altered the apparent lightness and glossiness of subsequently viewed surfaces. This, Motoyoshi et al. proposed, suggested that a neural mechanism existed which was sensitive to such statistics of skewness.

This conclusion was shown to be flawed for a number of reasons. Landy (2007) published a response shortly after the original paper arguing that while these parameters of luminance histograms might be convenient mathematically, they did not correspond precisely to the computations used in perceptual judgements. Luminance histogram statistics are not the whole story for the perception of gloss or lightness, as a great deal also depends on the surrounding environment and surfaces. Perceived specular reflections such as highlights and pseudoimages are also necessary, and surroundings need a pattern of illumination consistent with statistics of natural scenes. Highlights must be positioned realistically, relative to the shading profile of the three dimensional surface. Glossy images may well have a skewed luminance histogram, but this is not a predictor of all images showing glossy objects - skew of the luminance histogram ignores all of the other cues (or pseudocues) accepted as being important to perceiving a glossy surface. Furthermore, Fleming (2014) made the point that this kind of diagnostic computation has the disadvantage of being fooled when the assumed statistics of the real world are violated; when in reality, gloss constancy is not flawed to this degree.

Anderson and Kim (2009) further criticised the proposals of Motoyoshi et al. by showing that the stimuli used in image analysis were not representative of the full range of possible stimuli encountered in the real world. The correlations only arose, they argue, because of the limited space of surface geometries, reflectance fields and illumination fields which Motoyoshi evaluated. The authors emphasised that photometric statistics fail to be predictive as they are void of any structural information required in distinguishing different types of surface attributes. The perception of gloss depends critically on consistency in location and orientation of highlights, relative to the shading profile and the three dimensional surface geometry; and this cannot be deduced from skew computations, as all information regarding location is discarded. To illustrate this point, Anderson and Kim made a number of images of glossy surfaces that had a negative luminance histogram skew. They also showed that Motoyoshi's adaptation experiment gave the same results for any level of luminance contrast, demonstrating that this was not exclusive to gloss perception. Any proposed statistic of this kind would have to be capable of reliably discriminating between different contributions to an image. In a second paper (Kim & Anderson, 2010) the adaptation experiment of Motoyoshi et al. was replicated, and no consistent after-effect was found. Adaptation to zero-skew adaptors produced after-effects similar to positively skewed adaptors, and negatively skewed adaptors produced no reliable after-effects. Wijntjes and Pont (2010) investigated whether Ho et al.'s findings (2008) of relief height correlating with perceived gloss could be explained by Motoyoshi et al.'s gloss predictor. However skewness of luminance could not account for this effect.

Ultimately, all attempts to devise a single diagnostic statistic - not directly related to any physical parameter that generates a gloss percept - for the perception of gloss have failed, as have attempts to characterise the entirety of perceptual gloss using a single proposed mechanism. Many studies have successfully characterised perceptual gloss to some extent, but none encapsulate the wide range of characteristics which affect perceived glossiness. It is not as yet fully understood why the visual system interprets interactions of shape, illumination and specularity in certain ways. Additional confirmation of these conclusions can be seen in several other studies: Ji, Pointer, Luo, and Dakin (2006) showed that visually scaled gloss data do not correlate with conventional glossmeter measurements over the entire range, demonstrating that the measurement of a single physical attribute is insufficient to account for perceptual gloss. Lindstrand (2005) also argued that the nature of perceptual gloss is too complex to be characterised by a single instrument. An example of this in practice can be found in the study by Nefs et al. (2006), investigating whether gloss influenced the perceived relief of a surface. Differences in illumination direction induced a change in perceived relief, but surprisingly, no systematic difference was found between matte and shiny surfaces. This seems to contradict the evidence discussed above. However, perceived gloss was assumed to be based entirely on specular highlights - therefore the 'surprising' findings were obtained as a result of neglecting to take multidimensionality into account.

1.5.14 In favour of a gestalt approach

Research on the perception of gloss has, to date, tended towards the conclusion that the visual system does not attempt to calculate or approximate the physical dimensions of surface reflectance or surface properties, but instead seems to analyse a constellation of cues and pseudocues in making these perceptual judgements. The sum of these object and scene cues forms tertiary properties of the perceived image. Fleming, Torralba, and Adelson initially voiced support for this approach in their 2004 paper investigating the power of specular reflections in perceiving the three dimensional shape of an object; since then, a great deal of evidence and support in favour of this approach has emerged.

In 2010, Wendt et al. showed that observers used several different kinds of information available in making judgements of gloss, to varying degrees (motion, disparity, and colour).

All types of cue investigated improved overall gloss constancy, both when used alone and in conjunction with other cues, but observers showed differences in their prioritisation of the various cues, when presented with multiple kinds of information. Leloup, Pointer, Dutré, and Hanselaer (2012) uncovered similar responses - observers were asked to make pairwise comparisons of real life stimuli, which incorporated multiple perceptual cues for glossiness. These comparisons were used to derive an overall scale of perceptual gloss. Differences in both distinctness of image and luminance affected perceived gloss. However, different strategies of evaluation were found between observers, as they attributed varying levels of importance to the different cues.

Moreover, cue (and pseudocue) selection differs from task to task for all observers. In a study investigating the cues used for comparative judgements of gloss, observers relied on whichever most reliably distinguished the pair of stimuli (Marlow and Anderson, 2013). Images differed in specular coverage, sharpness and contrast - so if there was high variability in specular coverage, but low variability in sharpness and contrast, gloss judgements would be strongly predicted by specular coverage. Marlow and Anderson concluded that in static images presented monocularly, judgements of perceptual gloss rely on a heuristic weighting of cues for the characteristics of specular reflections. However, for this particular set of images it must be remembered that while weighted combinations of the variables used strongly accounted for observers' perceptual judgements, this was for a limited set of surfaces under very specific conditions (Fleming, 2014).

It is evident that we can recognise the physical nature of objects from information available in the key features of the appearance of gloss (Fleming, Wiebel, & Gegenfurtner, 2013; Ged, Obein, Silvestri, Le Rohellec, & Viénot, 2010). There is collective agreement that the brain does not, and could not, perform computations of inverse optics, as there is not enough information available to the visual system to invert the process of image formation and arrive at the base surface and illumination properties (Anderson, 2011). Fleming supported these conclusions in a recent review paper (2014), and argued that findings regarding the orientations and position of highlights imply that the goal of perception is not an inverse optics approach or a 'bag of tricks' method, but rather that it aims to characterise the overall 'look' typical of particular surfaces, and how this appearance tends to vary. Constellations of low- and mid-level image measurements convey the extent to which the surface manifests specular reflections; and statistically informative appearance characteristics can be measured which indicate the nature of underlying changes in material properties. These can be correlated between samples of related materials, to establish the typical appearance of a glossy surface. Fleming also proposed that such 'statistical appearance models' are more expressive (as a result of treating the image as a gestalt), and easier to compute than the physical parameters; and are therefore a powerful mid-point between a 'bag of tricks' and inverse optics.

A mid-point model has a considerable advantage over the more extreme models, Fleming continues, in that it has the capability of predicting what new, unseen surfaces of similar properties might look like. This is more efficient than the long-division inverse optics method, and more accurate and reliable than depending on a standalone diagnostic image statistic. There is a general assumption that salient features are likely to relate in some systematic way to the underlying properties of the materials, and it seems that observers use the most salient (in terms of variation) perceptual cues when making judgements of relative gloss. Furthermore, Fleming rightly points out that the visual system does not necessarily care about representing the physical dimensions in a way true to their physical organisation. For instance, hue is perceptually circular, in that a perceptually valid colour wheel can be produced with reds and blues blending into one another sequentially through purple, whereas in physical terms, wavelengths are linearly organised and purple light can only be composed of a mixture of multiple wavelengths. We have therefore no reason to assume that the visual system makes use of an internal scale that is wholly true to the physical scales of dimension.

1.5.15 Summary

Initial theories of gloss perception relied on the use of a single dimension on a physical scale. This was soon refuted, and attention turned to a multidimensional approach, as interactions with 'unexpected scene variables' indicated that the perception of gloss was far more complex than initially thought (Ho et al., 2008). Some shifted to the other extreme and proposed a diagnostic image statistic, but this was quickly overturned on the grounds that the proposed statistic was flawed and that such a statistic would not necessarily be reliable. Discussion returned to the consensus that perceptual gloss is reliant on multiple dimensions. This carries the implicit assumption that a solvable formula exists for the multiple dimensions, given sufficient investigations; yet recent results indicate that this assumption too may be oversimplified. Not only is there variability between the salience of different features from object to object; there is also fluctuating inter-observer agreement about the applicability or salience of different perceptual cues; and differences in the importance attached to these cues and their salience between observers. As if this wasn't enough, the judgements made by observers in response to real life stimuli are not easily replicated in experimental simulations, and this suggests that we have yet to identify the full extent of relevant information used in veridical perceptual judgements. When there is limited information available from stimuli, observers are forced to prioritize the most salient distinguishing factor, which results in great inter-observer disagreement. However when there is a broad spectrum of perceptual cues and a richness of information not normally present in simulated images (when the images are as close as possible to achieving a real life experience) so that observers are not forced to prioritize the information available - then there is much greater consistency in responses. This suggests that more work is required to identify the additional perceptual cues on which observers rely, and the nature of their interactions with established cues.

Preface to Chapter 2

In the first experiment of this thesis, the fundamentals of translucence perception are addressed. While physical translucence is determined by the nature of light propagation (light scattering and absorption) through a material, as outlined in the Introduction it is uncertain how the visual system interprets the information available. This is the first study to address perceived translucence experimentally, in relation to the physical absorption and scattering of light.

Technical note on approaches to statistical model selection taken in this thesis:

When performing model comparisons there are a number of possible methods, including Aikake Information Criterion (AIC), Bayesian Information Criterion (BIC), and the comparison of least squares. The comparison of least squares assumes that residuals are normally distributed. The AIC and BIC both essentially provide an estimate of the goodness of fit of a model using log likelihoods, weighted against the number of hypothetical parameters. The BIC weights more heavily against more complex models compared to the AIC and so tends to select more parsimonious models. Here, we have used nested hypothesis tests with log likelihood ratios. It is argued that the nested hypothesis method is the clearest and most straightforward method for determining which models provide the best fit in our experiments. The nested hypothesis test compares the different models as pairs and only selects a more complex model when it explains the data significantly more effectively than a simpler model. The inherent conservatism of the method towards simpler models achieves the same goals as AIC or BIC but with clarity and simplicity.

AICs, while indicating the amount of variance still not accounted for by the data, might give a lower AIC value for a particular model when it is not clear that it fits the data significantly better than another model. They can be more complicated to interpret because of this, as one can look at both the absolute values of the AICs and the statistical tests of the differences for AICs for different models, which may have discrepancies. We applied AICs to the interpretation of study 1, and found no difference of conclusion to the nested hypothesis approach.

Technical note on image rendering:

Throughout this thesis, physically-based image rendering has been used to create stimuli. When tone mapping is used to convert images for display on standard monitors, this can compress high luminance values, which would have an effect on perceptual judgements - particularly perceptual judgements of gloss. For all experiments, we used linear compressions so as not to compress the high luminance values.

Chapter 2

Beyond scattering and absorption: Perceptual un-mixing of translucent liquids

2.1 Summary

Identification of many natural qualities, from the ripeness of fruit to the fitness of potential mates, depends upon translucence in addition to colour or texture (Fleming & Bülthoff, 2005). We ask whether perception of translucence is based on estimations of scattering and absorption of light, or whether observers use statistical pseudocues associated with specific familiar materials. From images of milky tea, observers judged tea concentrations regardless of milk concentrations, or vice versa. Stimuli were photographs of real tea or photorealistic physically-accurate rendered images. In real tea, absorption and scattering interact as the strength of tea and milkiness vary. In the rendered stimuli there was no interaction. If judgements encapsulated the tea-specific interactions between milk and tea on scattering and absorption, performance would be best for real stimuli. If judgements rested on independent estimates of light transport properties in the constituents of the mixture, with no interaction, performance would be best with renders. Estimates of the contributions of each physical component to perceived milkiness or tea strength were made

using maximum likelihood conjoint measurement (Ho et al., 2008). Separability of the two physical dimensions was better for real than rendered teas, but neither was perfect. Simulated decisions based on image statistics show that milkiness judgements were predicted by colour saturation and strength judgements were predicted by a weighted sum of lightness and spatial gradients of saturation, for both real and rendered stimuli. An explanation of observers' behaviour through pseudocues suggests that perceptual un-mixing depends on ad hoc perceptual models, rather than estimates of underlying material-light interactions.

2.2 Results and Discussion

Many natural materials not only reflect light from their surfaces but are also translucent. Translucency is an optical characteristic which is caused by light scattering below the surface of an object, that is, it is scattered within the material. The two most important physical parameters that determine how light is transported within a material are the scattering and absorption coefficients, which essentially capture the rate at which light spreads and is attenuated as it travels through the material. Making judgements of the purity and concentration of mixtures of translucent materials (e.g. the strength of tea or its milkiness) is not simply a matter of identifying a material as being translucent, but also of estimating its scattering and absorption parameters. Very little is currently known about how we perceive translucence (Anderson, 2011). The results we present here provide the first systematic assessment of how humans perceive scattering and absorption of light within translucent materials. We do so by asking whether observers can perceptually un-mix constituents that primarily scatter or absorb light in a mixture. We use mixtures of milk and black tea (milk scatters light while tea absorbs light - Aernouts et al., 2015; Narasimhan et al., 2006 - but they do interact slightly, Hasni et al. 2011) and we assess un-mixing by measuring observers' abilities to make independent perceptual estimates of the concentrations of the constituents. The method of maximum likelihood conjoint measurement (MLCM: Ho et al., 2008; Knoblauch & Maloney, 2012) allows us to estimate the contributions made by two distinct physical variables to a single perceptual judgement; the influence of both factors on the judgement can be quantitatively characterised. Comparative judgements of a set of photographs of real glasses of tea (Figure 2.2a), which varied in two physical dimensions (milkiness and strength of tea), were made for judgements of milkiness and of tea strength in separate tasks. MLCM analysis then estimated the actual contribution of the two physical variables to each perceptual estimate, with independent, additive, or saturated combinations of the two physical dimensions potentially modelling the perceptual estimate. An independent model assumes that only one variable contributes to each perceptual estimate, an additive model allows for a simple additive combination of contributions from the two variables, and a saturated model allows complex interactions between the two variables. Figure 2.1 illustrates how the plotted responses of observers may look when giving a best fit to the three models. For each model, we computed goodness of fit to the data as log likelihoods. A nested hypothesis test was performed on successive fits to determine the most parsimonious model for the data by calculating whether an increase in the number of parameters in the model explained a significantly higher proportion of the variance. Prior to MLCM testing the stimulus sets were prepared so as to ensure that steps along the two physical dimensions were approximately equally discriminable (by means of maximum likelihood difference scaling - MLDS; Knoblauch and Maloney 2012 - to obtain stimuli that were equally spaced in units of d-prime). This enables us to conclude that any interactions found in the main MLCM experiment are not due to scale differences in discriminability in the stimulus sets.

2.2.1 Results from photographs of real tea

The independent model was rejected for all of the eight observers in the perceived tea strength task, demonstrating that both physical variables always contributed to perceptual judgements in this task. The independent model was rejected for seven of the eight observers in the perceived milkiness task. For most observers, the physical variables contributed in additive combination to perceptual judgements, meaning that every level of the distractor variable contributed a fixed offset to the perceptual judgement (Figure 2.3a and b). Only two of the eight observers needed a model more complex than the additive one to fit judgements of tea strength, and only one needed more than an additive model to fit judgements of milkiness. For perceived tea strength, two kinds of observers

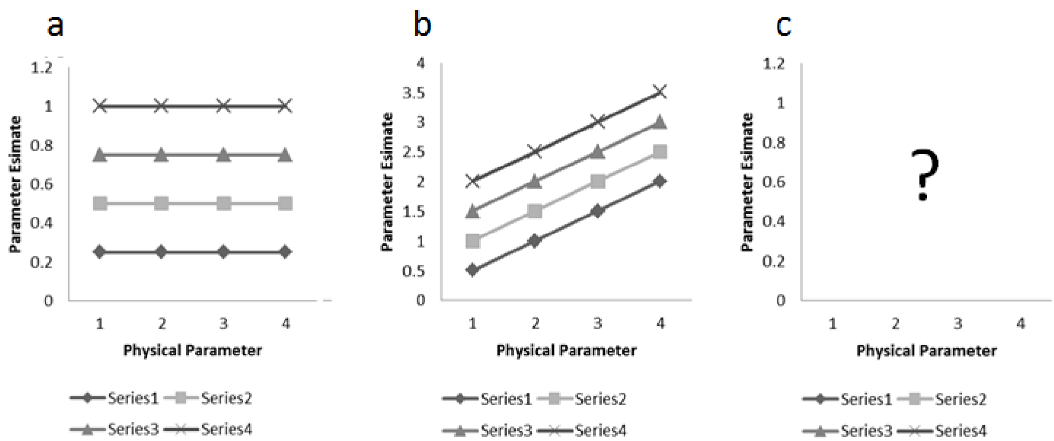

Figure 2.1: Parameter estimates (in d', on the y-axis) are shown as a function of two physical variables; levels of one physical variable are represented on the x-axis, and levels of the second variable are represented in the four plotted lines. a) illustrates a hypothetical independent model, where judgements of the parameters are assumed to be completely independent of one another - judgements would not be affected if the second variable were removed. b) shows a hypothetical additive model, which assumes that the second variable produces a simple additive confound. To predict the parameter estimate, the specific combination of variables would not be needed - just the variable under consideration, in order to be able to 'add' the right amount to the parameter estimate. c) For a saturated model, the full model would be needed to predict parameter estimates. There is no assumption of linearity, and complex interactions are allowed.

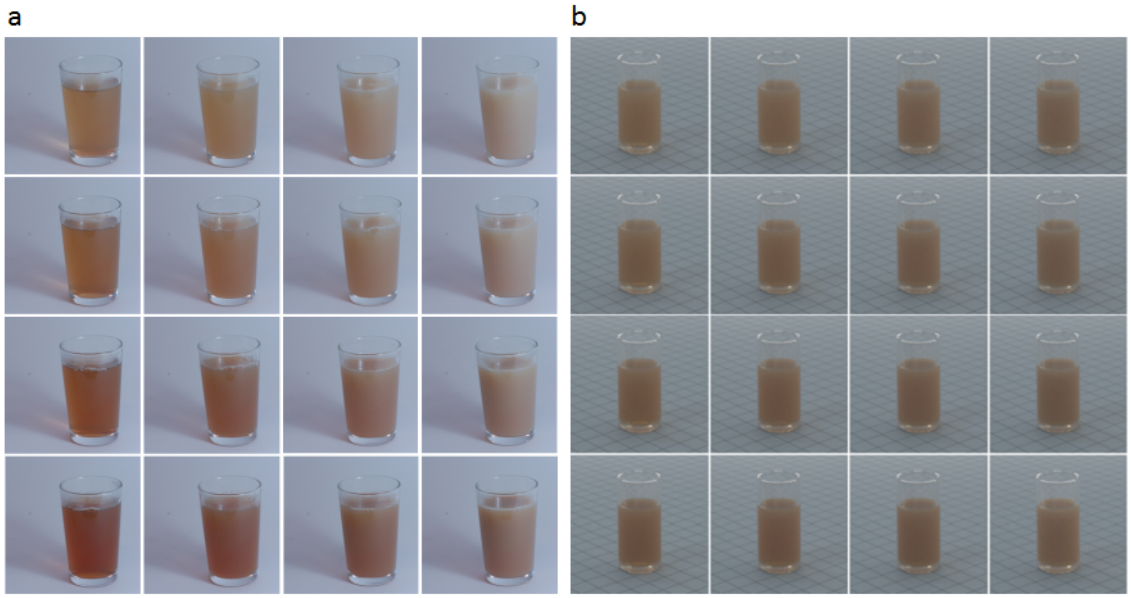

Figure 2.2: (a) The set of 16 real stimuli used in the MLCM task, with milkiness increasing left to right and tea strength increasing top to bottom. b) The set of 16 rendered stimuli used in the MLCM, with simulated 'milkiness' (scattering) increasing left to right and simulated 'tea strength' (absorption) increasing top to bottom.

were evident - for some, increasing milkiness increased the perceived tea strength, whereas for others it decreased perceived strength. In the milkiness task, increased tea strength consistently decreased the perceived milkiness, with the exception of participant BC. Participants' judgements were largely driven by the physical parameter that we asked them to judge, as regressions of the additive models showed greater contribution of 'relevant' variables towards parameter estimates (as seen in the additive plots in Figures 2.3a and b).

As required for reliable perception of tea mixtures, we found that whenever concentrations of either tea or milk increased so did perceptual estimates of their concentration. However, un-mixing was not perfect: additive rather than independent models dominated the results, indicating consistent additive effects of milkiness on judgements of strength, and vice versa. The magnitude of the additive contribution was small but significant across the discriminable scale. Participants are therefore far from perfect at un-mixing real tea and milk but are responding to tea and milk concentrations in a systematic manner. What then, exactly, are observers doing?

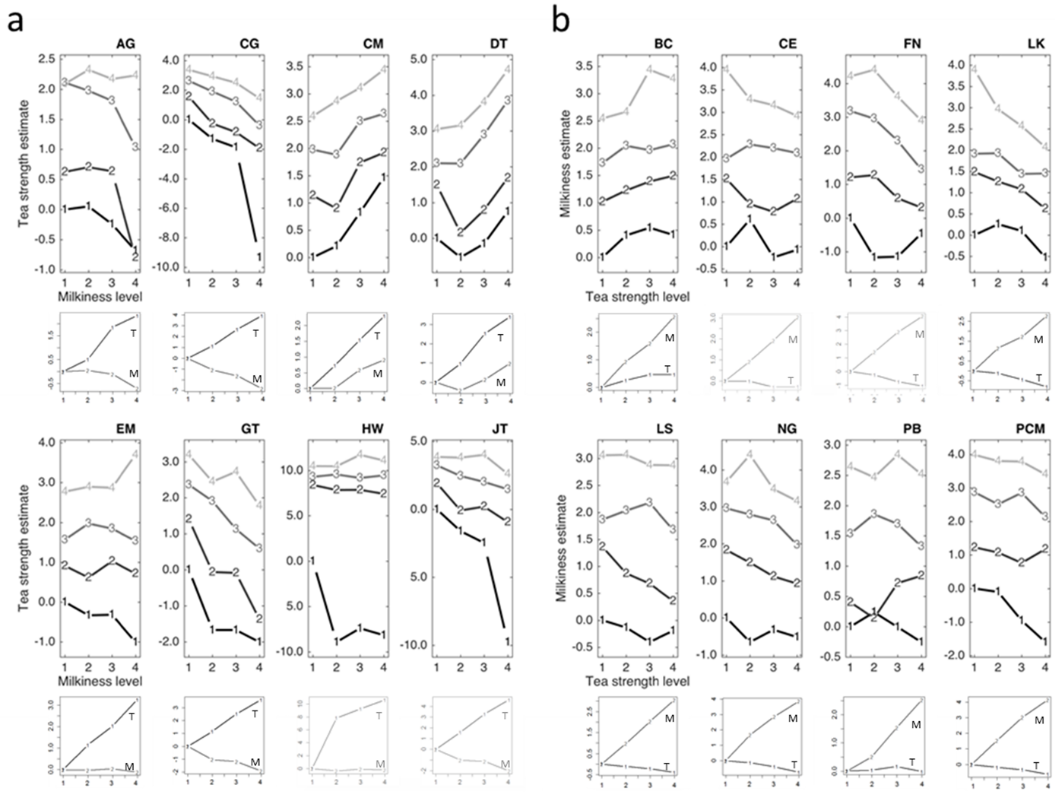

Figure 2.3: Upper plots with four lines: a) the best-fitting saturated model of perceived tea strength estimates for each individual (in units of d', on the y axis) as a function of physical milkiness (x axis). Numbers 1-4 within the plots denote low to high levels of physical tea strength. b) The best-fitting saturated model of perceived milkiness estimates for each individual (in units of d', on the y axis) as a function of physical tea strength (x axis). Numbers 1-4 within the plots denote low to high levels of physical milkiness. Lower plots with two lines: a) the best-fitting additive model of perceived tea strength (y axis) as a function of level of physical variable (x axis). b) The best-fitting additive model of perceived milkiness (y axis) as a function of level of physical variable (x axis). In these plots the line labelled 'M' denotes the contribution from physical milkiness and the line labelled 'T' denotes the contribution from physical tea strength. For results where the additive model provided the best fit, the additive graph is in bold.

The information available for visual perception of the material properties of objects derives from the physical properties of a material, which in turn determine the nature of its interaction with light. Real materials exhibit complex combinations of a variety of types of physical light transport, such as absorption and scattering, operating over multiple spatial scales. In translucent materials, light is scattered not just from the surface of the material but also within the volume. In tea, specifically, the particles of predominantly light-absorbing tea agglomerate onto fat globules in predominantly scattering milk, so, as milk concentration increases the absorbing power of a given concentration of tea decreases (Hasni et al., 2011).

Perhaps the way real milk and tea interact in their effects on scattering and absorption interfered with un-mixing. If our perceptual judgements of mixtures were based purely on estimates of scattering and absorption of the separate constituents without taking proper account of the interactions between milk and tea that affect scattering and absorption in real tea (Hasni et al., 2011), then images of artificial computer-rendered 'tea' in which 'milkiness' affected only scattering and 'tea-strength' affected only absorption, without complex interaction, should be easier to un-mix.

Furthermore, perceptual judgements with the two classes of stimuli, real and rendered, that arise from subtly different physical processes and as a result contain subtly different image statistics, allow us to test the extent to which particular pseudocues (presumably learnt with real stimuli) can explain judgements.

2.2.2 Results from rendered stimuli

The stimuli (Figure 2.2b) were created with a physically-based computer ray-tracing renderer. They were highly realistic and observers thought that they were images of real glasses of tea. We asked observers exactly the same questions as in Experiment 1, that is, to judge 'tea strength' or 'milkiness'. Whilst observers felt they could estimate each perceptual property separately and thought the two to be conceptually distinct, just as in Experiment 1, the perceptual estimates we extracted with MLCM were described by complex saturated combinations of both physical properties, for the majority of participants (see Figure 3a and b). For tea strength, four out of five were best fit by a saturated model, and for milkiness, three out of five were best fit by a saturated model, with an additive model fitting the rest. The independent model never provided an optimal fit. In perceptual terms, variation in scattering and absorption did not map separably onto perceived milkiness and strength. Observers cannot therefore be basing their judgements on independent estimates of the light transport properties of absorption and scattering.

2.2.3 Differences between results for real and rendered stimuli

How can it be that perceptual un-mixing of real tea was so much more successful than the un-mixing of our artificial tea? For real tea, judgements departed from the true concentrations only by an additive contribution from the distracting variable, whereas judgements with rendered tea required modelling full interactions between the two physical variables. Two potential artefacts that might explain these results need to be ruled out. First, the real tea stimuli show a greater degree of visual variation, including bubbles at the surface, than the rendered stimuli. It is not clear, however, that this variation should make the un-mixing task easier. If anything it should introduce further interference. Moreover, the two stimulus sets elicited similar ranges of perceptual difference, as measured by the d-prime measure of discriminability in the preliminary MLDS task, and shown in the extracted perceptual estimates shown in the MLCM graphs in Figures 2.3 and 2.4. Statistical power might also be an issue, as the number of trials differed between experiments. We reanalysed the data, sub-sampling the same number of trials for each, and found that there were no meaningful changes to our findings.

It is true that there are more differences between the two sets of stimuli than just the ways in which the liquids have been created. The lighting appears more direct in the real stimuli, the stimuli levels are not quite the same, and the shapes of the volumes are also slightly different. In the printed versions of these stimuli, a bright tea coloured patch is visible within the 'shadow' of the real stimuli where light has travelled through the volume; this effect is also present (although more subtly) in the rendered set when presented on a high-quality screen. It is very hard to match the exact levels of real and rendered images, so the two sets of stimuli levels used were selected for producing a linear range in MLDS testing; this extent of level-matching was the most achievable that could be attained. There are clearly differences between the two sets of stimuli, and so it is an open question whether those differences might affect the results obtained, however the fact

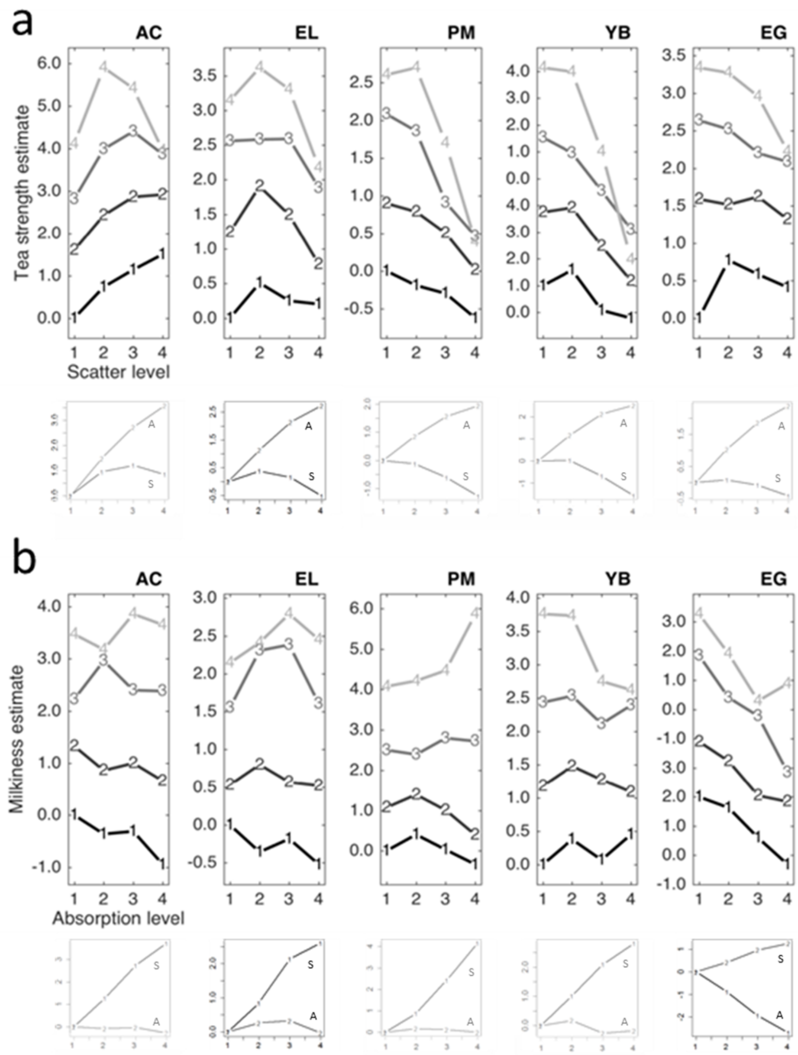

Figure 2.4: Upper plots with four lines: a) Perceived tea strength estimates for each individual (in units of d', on the y axis) as a function of physical scatter (x axis). Numbers 1-4 within the plots denote low to high levels of physical absorption. b) Perceived milkiness estimates for each individual (in units of d', on the y axis) as a function of physical absorption (x axis). Numbers 1-4 within the plots denote low to high levels of physical scatter. Lower plots with two lines: a) the best-fitting additive model of perceived tea strength (y axis) as a function of level of physical variable (x axis). b) The best-fitting additive model of perceived milkiness (y axis) as a function of level of physical variable (x axis). In these plots the line labelled 'S' denotes the contribution from physical scatter and the line labelled 'A' denotes the contribution from physical absorption. For results where the additive model provided the best fit, the additive graph is in bold.

that the same question could be asked of different stimuli and the resulting patterns of responses could be fit to the same simple and complex image statistics suggests that those effects would not be great. It could be argued that the responses we have obtained are from different parts of the same 'tea-space', but the fact that the differences between the patterns of response are not huge means we are roughly in the same region of tea-space. Due to the extent of the comparability between the interpretations of results by modelling responses using complex image statistics, it is very unlikely that the differences between experiments was entirely a result of alternate lighting or shapes of glasses.

2.2.4 Parameter estimates based on pseudocues

Perhaps, rather than estimating light transport properties and, from them, inferring material properties, observers use pseudocues in the image that are direct heuristic estimates of concentrations of specific materials (Chadwick & Kentridge, 2015). By this reasoning, observers are better at perceptual un-mixing when physical dimensions correspond to previously encountered materials, whereas applying learnt pseudocues to our rendered images would be expected to produce less effective performance. The data suggest that this is the case. With rendered stimuli, complex interacting combinations of the physical dimensions contribute to perceptual judgements of strength. These can produce non-monotonic effects - sometimes, increasing scatter (adding more 'milk') makes the liquid look stronger, and sometimes less.

2.2.5 Image statistics as pseudocues

If observers are not making estimates of light transport properties, how are they un-mixing the real materials? We conducted a number of simulations using image statistics of the tea-region of the images to investigate potential pseudocues upon which observers may have based their judgements (see Figure 2.5). We used a statistic from each image to simulate decisions in the MCLM task (for a pair of images, the simulated decision labelled the image with the larger value of the statistic 'milkier' or 'stronger'). We then calculated the goodness of fit between simulation and data for estimates of the contributions of the physical parameters of the tea to perceptual judgements in a full, saturated, MLCM model. Mean colour saturation provided a good explanation of performance in the milkiness task, for both real and rendered stimuli (adjusted r^2 value = 0.920 and 0.970 respectively). No simple statistics provided a good account of behaviour in the tea strength task. The inclusion of spatial information was crucial for getting a good fit to the data obtained with observers for this task. A linear mixture of brightness and colour saturation gradient (from the top surface of the liquid into the tea volume, summarised by a fitted exponent describing the space constant of variation in saturation as light penetrates the volume) provided a good account for both the real and rendered stimuli (adjusted r^2 value = 0.812 and 0.894 respectively).

The similarities of statistical model predictors between the real and rendered images suggest that there is a level of consistency in how observers responded in making perceptual judgements of the real and rendered images. They may also go some way towards characterising the observers' strategies when making perceptual judgements. The spatial distribution of information within an image is well-known to be vital for making judgements of material properties (see Landy 2007 for an example). While it is understandable that the image statistics we tested correlate to some extent with the perceptual judgements (and indeed levels of milkiness and tea strength, or physical scatter and absorption) these statistics don't completely capture the patterns of perceptual decisions seen in both the real and the rendered experiments. The non-monotonicities in the obtained data with the rendered stimuli suggests that there was something wrong with the information used by observers when making judgements about the rendered stimuli at low levels of scattering, and therefore that the pseudocue used was inappropriate for the rendered stimuli.

We can conclude that it is entirely possible for a single perceptual model to produce very different responses for what appear to be very similar stimuli, as we found in the comparison of real and rendered stimuli. The discovery of image statistics that can approximate real observers' responses in both of the experiments illustrates that there may well be a 'short-cut' to achieve modeling of a material substance. This short-cut takes the more complex interactions of the volume into account without necessarily estimating the physical variables of light transport, since it is possible to find pseudocues that implicitly capture the effects that the physical dimensions have on the perceptual dimensions. The statistics of best fit identified here are not intended to answer to the question of how observers perceive perceptual invariants of aspects of translucence. Instead they demonstrate that a single

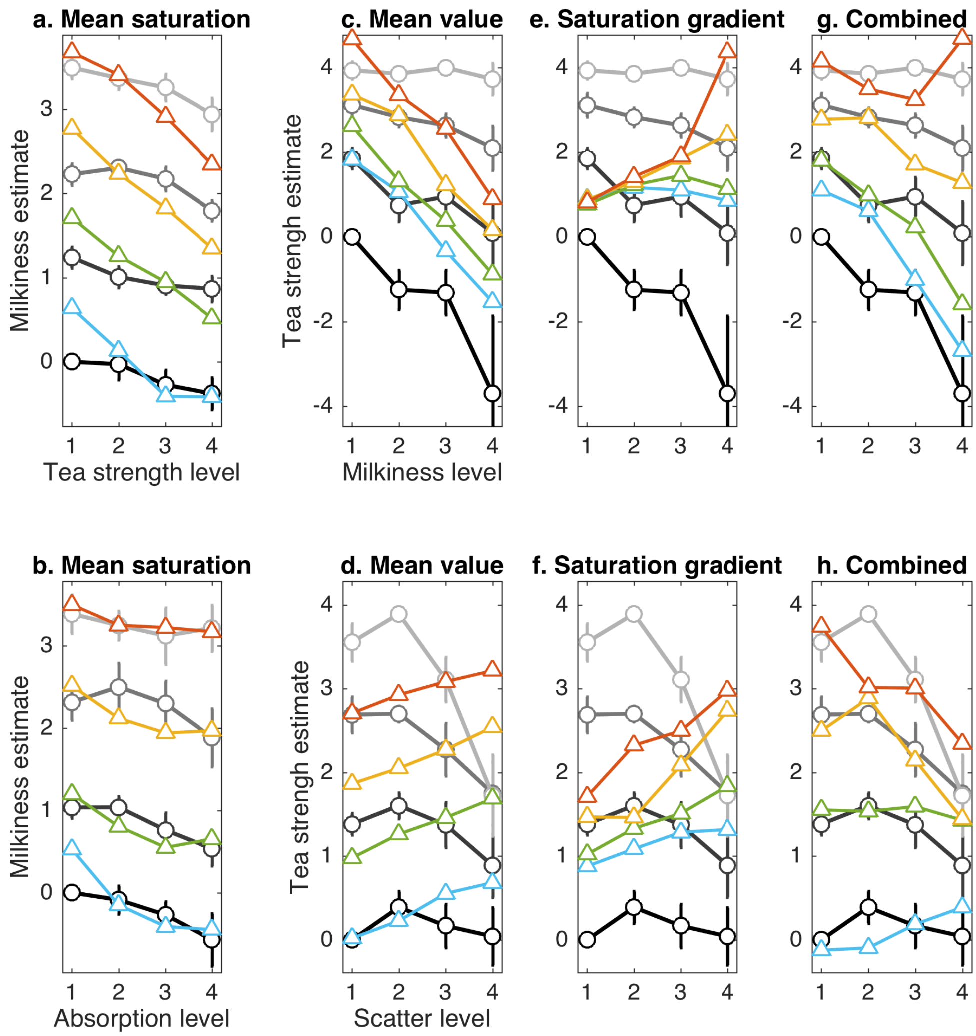

Figure 2.5: Average milkiness and strength estimates (greyscale lines and symbols, with 95% confidence intervals across observers) for real and rendered stimuli top and bottom rows respectively), with accompanying model fits (coloured lines and symbols) from an ideal observer whose responses are governed by candidate image statistics (see text for full details). a) & b) Milkiness estimates with ideal observer responses based on the mean of saturation, adj. $r^2 = 0.920$ and 0.970 for real and rendered. c) & d) Strength estimates with ideal observer responses based on the mean of value, adj. $r^2 = 0.535$ and 0.670. e) & f) Strength estimates with ideal observer responses based on the space constant of gradients of saturation, adj. $r^2 = 0.096$ and 0.207 for real and rendered. g) & h) Strength estimates with ideal observer responses based on a weighted sum of the mean of value and gradients of saturation, adj. $r^2 = 0.812$ and 0.894 for real and rendered.

statistical model can optimally describe observers' behaviour for judgements of both real and rendered stimuli. Critically, this demonstrates that a single common mechanism can potentially account for responses to the two stimulus sets, despite the marked differences in observers' patterns of behaviour. We identify an effective statistical pseudocue for describing the behaviour of observers making a specific judgment about one set of stimuli and show that when observers are asked to make the same judgment about a different set of stimuli the same statistical pseudocue still describes their behaviour despite radical differences in response patterns between the two stimulus sets. We have identified potential pseudo-cues to translucence properties that operate across different types of materials - the real and rendered teas. One would, of course, expect that additional processes would come into play in producing estimates of those properties that remained consistent across contextual changes (e.g. in lighting, object shape etc.), although it is also possible that our perceptions of translucence sometimes only show weak constancy across contexts.

2.2.6 Individual differences

There are individual differences evident in the MLCM analyses of both experiments producing patterns of response that appear qualitatively different. It is worth mentioning, however, that individual differences within each task or experiment are not as large as the differences between the experiments. It is not surprising that we find these individual differences. If a trial is ambiguous (as indeed many were, as the tasks were designed to challenge observers) and observers cannot find consistent cues, then large differences will be inevitable as observers try out different strategies to make judgements about difficult comparisons.

2.2.7 Conclusion

These results have implications for the fundamentals of translucence perception. We can of course perceive translucence; it is clear, however, that we do not independently perceive the physical determinants of the composition of light reaching the eye, in that we do not perceive translucent substances as simple linear mixtures of the scatter and absorption properties of the constituents. Fleming and Bülthoff (2005) also examined the translucence properties of volumetric materials and our perception of them. Although a wide number of factors (including highlights, contrast, colour and blur) can all contribute to perception of translucence, they agreed that the visual system is not using them to compute estimates of light transport properties as an intermediate step in material perception. Instead they suggested that a generative model parses scenes into key regions and allows degrees of variation to be perceived without requiring knowledge of physical light transport factors. Our results support this suggestion. Using real and rendered stimuli has allowed us to generate conditions that make explicit, separable predictions for direct estimation of light transport vs reliance on pseudocues and so test the viability of Fleming and Bülthoff's approach. Pseudocues are, in a sense, indirect estimates of the effects of physical light transport, and could even account for complex combinations of distinct types of information. They are, however, by no means driven directly by light transport characteristics. This is perhaps more impressive or useful than calculating the physical properties of light, as the visual system appears capable of accounting for complex physical interactions in material substances when determining the best proxies to use in making perceptual decisions.

It is possible that our observers were using pseudocues based on the means of saturation and brightness of the tea-region of the image and on spatial gradients of saturation, since an ideal observer that bases decisions on these statistics is capable of distinguishing the images in a way that mimics responses of real participants. The statistics we really use are probably more complex, as our simulation doesn't capture the full range of responses by observers. To confirm the use of pseudocues with a greater degree of certainty, we would need to identify and manipulate a range of pseudocues and determine their effect on parameter estimates, whilst precisely monitoring physical scattering and absorption.

More generally, these results provide evidence in favour of a general theory of vision which denies that we see the physical determinants of volumetric or surface properties. The visual system does not 'know' the physical laws of light scattering and absorption, and does not perform calculations of inverse optics to estimate these properties.

2.3 Supplemental Information

2.3.1 Experimental procedures

MLDS

Maximum likelihood difference scaling was used to estimate how changes in just one of the two physical dimensions translated to perceived difference estimates. Two pairs of stimuli differing along one physical dimension were presented simultaneously, and observers asked to decide which of the two pairs showed a greater perceptual difference. Quadruples (the two pairs) were chosen in accordance with the methods of Knoblauch and Maloney (2012): only non-overlapping quadruples were used (that is, for the four values, a, b, c, d, of the parameter being manipulated, $a<b<c<d$). The maximum likelihood perceptual scale was computed using the Knoblauch and Maloney MLDS package for R. When estimating a scale, the value of the second 'irrelevant' physical dimension was fixed at a low level to avoid interference. As the stimuli used were familiar substances, observers could be asked to report the greatest difference in 'milkiness' or 'strength' of tea, rather than referring to the physical variables in the instructions for participants. This procedure allowed us to calculate estimates of perceptually constant scales for both real and rendered stimuli. These scales produced a perceptually linear 'tea-space' for use with maximum likelihood conjoint measurement, to ensure that any interactions in the results of MLCM analysis would be the result of genuine interactions between variables rather than non-linearities in perceptual scaling.

MLCM

Maximum likelihood conjoint measurement aims to determine how two different physical properties contribute to a single perceptual dimension of judgement, and the extent to which they interact with one another. Pairs of stimuli were presented to observers while manipulating both physical variables from trial to trial. Observers were instructed to make a judgement about only one perceptual dimension per task, reporting which of the pair corresponded to 'more milky' or 'stronger' tea.

Data were analysed with the Knoblauch and Maloney MLCM package for R to choose between three different ways of modeling the data: independent, additive, and saturated.

In an independent model, perceptual judgements are assumed to depend on just one of the two physical properties and the model therefore fits three hypothetical parameters (one for each value of the physical variable used, except the first level as this is fixed). An additive model assumes that both physical variables contribute in a simple compound manner, for example increases in absorption and scattering might both increase the milkiness estimate but a particular amount of absorption would have exactly the same effect for all amounts of scattering. The model therefore fits six hypothetical parameters - one for each level of each physical variable, except the first levels of each variable. A saturated model makes no assumption of linearity and allows for complex non-linear interactions so that the effect of one physical variable depends on the value of the other physical variable. This model needs separate parameters for every combination of the levels of the physical variables except the combined first levels of each variable, in this case fitting fifteen parameters.

When testing the goodness-of-fit of the models, we used the log likelihood values in nested hypothesis tests to establish whether there were significant differences between the amounts of variance explained by each model. We interpret the results by starting with a comparison of the saturated model with the additive model, to see if adding parameters yields significantly better model fit. If this is not the case then we go on to test the additive model against the less complex independent model, again determining whether adding parameters produces a significantly better fit. Where these tests are inconclusive in identifying a model of best fit, we follow this up with a comparison of the saturated and the independent model. These nested hypothesis tests are based on a chi-squared approximation of the distribution of log likelihood ratios (Boes, Graybill, & Mood, 1974).

Statistical software

All calculations were performed using Knoblauch & Maloney's (2012) MLDS and MLCM packages for R (v2.11.1).

Methods for real stimuli experiment

To create controlled images of real tea, a 'master' tea solution was made with freshly boiled water before adjusting the strength by watering down as required. This volume was kept at a constant temperature to prevent the tannins precipitating and making the volume cloudy, as it was found that there was a just noticeable difference in the spectral composition of the light reflected from a glass of the tea solution detected by a spectroradiometer (Mahy, Eycken, & Oosterlinck, 1994) over a period of fifteen minutes as the liquid cooled. As several master volumes of solution were required to make sufficient stimuli for extensive piloting, a procedure was developed to create the optimal strength of master solution. Ten teabags (Tetley) were added to three litres of boiled water for three minutes. The same brew of master solution was used for all stimuli in the MLCM experiment. The total volume of liquid in each glass was 70ml, and the different stimuli were created by varying the amounts of tea solution, water, and milk (semi-skimmed, 1.8% fat).

An initial set of 20 stimuli was produced: 10 varying in milkiness, at the lowest strength of tea, and 10 varying in tea strength at the lowest level of milkiness. After MLDS piloting with multiple participants, an approximately linear perceptual scale was identified for each parameter within the range of stimuli. Four levels were identified on the perceptual scale for each of the physical variables with approximately equal perceptual differences, and the corresponding physical measurements were extracted. A new set of perceptually constant stimuli was generated with these values, with 16 stimuli in total (see Table 2.1 for the values in ml of each liquid used to create the final set of volumes, and Figure 2.2 for the set of photographic stimuli).

Table 2.1: The set of values, in ml, used to create the volume at each level of tea strength and milkiness

	Level of milkiness											
	1			2			3			4		
Level of tea strength	T	W	M	T	W	M	T	W	M	T	W	M
1	3.6	56	0.1	3.6	56	0.65	3.6	55	1.8	3.6	52	4.2
2	4.9	55	0.1	4.9	54	0.65	4.9	53	1.8	4.9	51	4.2
3	8.5	51	0.1	8.5	51	0.65	8.5	50	1.8	8.5	47	4.2
4	12.7	47	0.1	12.7	47	0.65	12.7	46	1.8	12.7	43	4.2

T = master tea solution, W = boiled water, M = milk in ml

Photographing the stimuli

Stimuli were photographed using a calibrated Nikon D80 camera which had been tested to ensure that no automatic lighting compensations were active in manual mode. Glasses of liquid were positioned against a white infinity-curve backdrop. The scene was lit by an overhead halogen lamp and a single fluorescent desk lamp with daylight spectrum, positioned to ensure some light passed through the volume of liquid towards the camera and also to create a visible shadow of the glass.

Observers

For the MLDS task, extensive initial piloting produced consistent results across several observers. One participant took part in the final MLDS experiment. For the MLCM tasks a total of sixteen observers took part, with eight completing each condition (either an estimated milkiness task, or estimated tea strength task). All participants were aged 18-25 and had normal or corrected to normal vision.

Apparatus

Experimental software was written in Matlab. Stimuli were presented on a calibrated NEC 2070SB CRT display (1064x768 pixels with refresh rate of 100Hz), controlled by means of a CRS Visage (Cambridge Research Systems). Observers were seated approximately 100cm in front of the screen in a blacked-out cubicle, and were asked to use a chin rest. Responses were made using a multi-button input device (Cedrus).

MLDS procedure

In a single trial, a fixation point was presented for 0.6 seconds, followed by a blank screen for 0.4 seconds, after which a quadruple was presented on the display screen for 3 seconds. A total of 126 trials were presented in each block, and repeated for 3 blocks. Observers were asked to judge whether the pair on the top or the bottom of the screen had a greater perceptual difference (of either milkiness or tea strength), and to indicate their decision by pressing one of two keys. The next trial was then triggered.

MLCM procedure

Participants were randomly allocated to one of the two MLCM conditions, milkiness judgement or strength judgement. A fixation point (0.4 seconds) indicated the start of a new trial, and pairs of stimuli were presented sequentially on the screen for 1.5 seconds each, separated by a blank screen for 0.5 seconds. Participants were asked to decide which of the two appeared to be either more milky or stronger. The program waited for a response to be given by the participant before moving on to the next trial. While the response time given was unlimited to prevent rushing a judgement, participants were told to respond as quickly as possible once they had made a decision. 136 trials were presented in each block, for 3 blocks, to give a total of 408 trials.

2.3.2 Methods for rendered stimuli experiment

Stimuli

Stimuli were created by simulating a glass tumbler, volume of liquid and scene in Blender, an open source 3D computer graphics program which can model 3D scenes and objects. Images were then rendered using the LuxRender ray tracing renderer. LuxRender simulates physical properties of materials, including their light-transmitting and light-scattering properties. It is based on PBRT (Physically Based Ray Tracing, Pharr and Humphreys 2004), and simulates the propagation of light through the scene in a physically realistic way (according to physical equations - for instance, through translucent materials).

The original stimulus set comprised 625 rendered images. The properties of the liquid were varied by manipulating the degree of physical light absorption and physical light scattering (where these manipulations were spectrally biased, such that at higher levels of absorption the liquid appeared brown and at higher levels of scattering the liquid appeared milky). A real-world lighting image probe employing natural spectral light distributions illuminated the scene. The liquids were produced by using appropriate equations to mix scattering and absorption parameters from two 'master' liquids, a 'strong black tea' with high absorption but no scattering (absorption coefficients 0.617, 1.886, 4.292, and scattering coefficients of 0, 0, 0 in LuxRender) and 'milk' with high scattering but no absorption (absorption coefficients 0, 0, 0, and scatter coefficients 9.5, 9.55, 10 in LuxRender). To create an image of 0.4 absorption and 0.5 scatter, the liquid would be a mixture of 40% 'strong tea' and 50% 'milk'. When the two properties were combined to simulate the volume, a weighted sum of the properties was produced, and the remaining volume was defined as a non-absorbing and non-scattering substance (that is, any part of the volume that was not defined by the scattering and absorption parameters described above essentially simulated the water - with no significant scattering or absorption of light).

There were 24 levels of physical absorption and 19 levels of physical scattering within the initial stimulus set. For MLDS testing, this was reduced to 11 levels of each variable ranging from 0.2-0.7 in increments of 0.05. The MLDS results were used to derive a perceptual scale in which differences in physical levels produced approximately equal differences in perceptual estimates. Following MLDS testing, the experimental set of stimuli was reduced to four levels for each parameter - the values 0.3, 0.35, 0.4 and 0.45 were used for physical absorption, and 0.25, 0.35, 0.45 and 0.55 for physical scattering (see Figure 2.2). As stimuli were presented in pairs in the MLCM experiment, the optimum task difficulty was determined on the basis of extensive piloting.

Apparatus

The apparatus used was the same as for Experiment 1. Observers were seated approximately 50cm away from the screen in a blacked-out cubicle, and were not required to use a chin rest, but asked to sit at a comfortable distance from the screen whilst maintaining visual acuity.

Observers

Three observers participated in the both the MLDS experiment and the MLCM experiment, with an additional two observers participating in the MLCM experiment. All were aged 18-25 and had normal or corrected-to-normal vision.

MLDS procedure

All observers completed both conditions of the MLDS task, on consecutive days to avoid fatigue. Completion of the two conditions was counterbalanced. The stimulus sequence was the same as for Experiment 1, with minor differences in timing. Instructions for observers were also the same as those for Experiment 1, with the exception that pairs of stimuli were presented on the left and right of the screen rather than top and bottom. A total of 330 combinations was presented in a randomised order, and repeated for 3 blocks, giving 990 trials in total.

MLCM procedure

All observers completed both conditions of the task on consecutive days to avoid fatigue. Completion of the two conditions by each observer was counterbalanced. Each condition consisted of 4 blocks of 180 trials, giving 720 in total. The stimulus sequence was the same as Experiment 1, with minor differences in timing, and pairs were shown side by side rather than sequentially.

2.3.3 Image statistics for both real and rendered images

Each image was converted from RBG to HSV values and the mean, standard deviation, skew and kurtosis was calculated for both the saturation and the value (where 'value' is akin to lightness or intensity in this context). We also extracted information about the saturation gradient extending downwards from the surface of the liquid, summarised by the space constant on a single exponential fit to the image data. Our model simulations assumed an ideal observer whose decisions in the MLCM task were determined by the relative values of image statistics. For example, the 'decision' as to whether one image was perceived as more milky than another would be based on which of the pair had a lower mean saturation. We chose to fit the model to the average estimates of milkiness and of strength, across participants. When calculating best fits of the ideal observer simulation to the real data we optimised a single scaling parameter on the image statistic in question since the scale of the real data (in d-prime units) depends in part on the amount of variability in decisions and there is no variability in extracted image statistics. Therefore, before averaging, the data shown in Figure 2.3 and 2.4 were normalised to each subject's maximum d-prime values, and the normalised average was re-scaled to the average d-prime range for that task. The averaged data are shown in Figure 2.5. To compare the ideal observer estimates with the averaged estimates from our participants, we used a simple linear model fit with an offset term and scaled contribution(s) from the image statistic(s) under test. The coefficients for all models presented here have $p<0.001$, apart from saturation gradient alone for the real and rendered strength judgements ($p=0.129$ and 0.043).

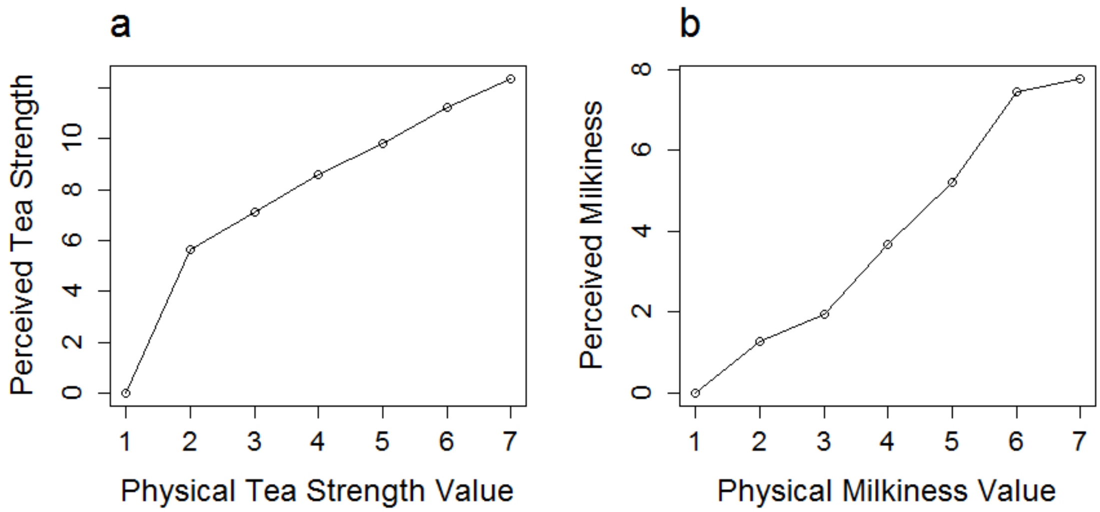

Figure 2.6: Related to Figure 2.2a: a) Perceived tea strength estimates (in units of d', on the y axis) as a function of level of physical tea strength (x axis). b) Perceived milkiness estimates (in units of d', on the y axis) as a function of level of physical milkiness (x axis), averaged across both participants and scaled to the average highest value. Values on the x axis denote levels of tea strength or milkiness within the real tea-space.

2.3.4 Experiment 1: Real stimuli MLDS results

For all observers, parameter estimates increased approximately linearly in relation to the physical parameter (ml of milk or of concentrated tea solution in a fixed volume - Figure 2.6), before perceived differences began to saturate. To obtain a physical scale of approximately equal perceptual steps we selected a desired d-prime range and used the MLDS data to look up physical values that would produce linear d-prime steps. We used four levels for each physical dimension, producing a perceptually uniform 'tea-space' of four-by-four stimuli. The strength values calculated for MLCM were approximately equivalent to levels 2, 3, 5, and halfway between levels 6 and 7 of the values used for MLDS. The milkiness values calculated were approximately equal to level 1, halfway between levels 3 and 4, and levels 5 and 6 of the values used for MLDS.

2.3.5 Experiment 2: Rendered stimuli MLDS results

Figure 2.7 shows parameter estimates for perceived strength and milkiness. Only values from the linear ranges of the two scales were used for the set of experimental stimuli. The four levels of physical absorption and scatter chosen for MLCM lie between levels 1 and 7

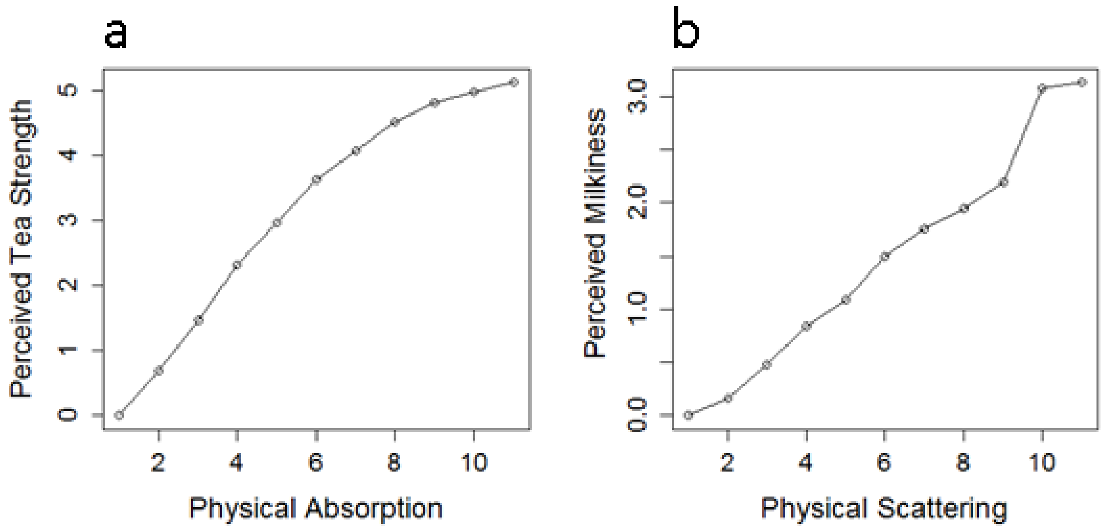

Figure 2.7: Related to Figure 2.2b: a) Perceived tea strength estimates (in units of d', on the y axis) as a function of physical absorption (x axis); normalised, averaged across all participants and scaled to the average highest value. b) perceived milkiness estimates (in units of d', on the y axis) as a function of physical scatter (x axis) normalised, averaged across all participants, and scaled to the average highest value. Values on the x axis denote levels of physical absorption or scatter within the simulated tea-space.

on the x-axes in Figures 2.7a and b.

2.3.6 Experiment 1: Real stimuli MLCM results

Table 2.2: Related to Figure 2.3a: Perceived tea strength task with real stimuli - log likelihood values for each participant, with nested hypothesis test p-values comparing the saturated model with the additive and the additive model with the independent.

	Observer							
Model/comparative test	AG	CG	CM	DT	BM	GT	HW	JT
1. Saturated model (15 parameters)	-208.753	-139.229	-228.900	-165.800	-183.739	-161.185	-125.473	-127.536
2. Additive model (6 parameters)	-218.272	-146.016	-232.333	-176.265	-193.703	-168.912	-144.856	-141.136
i. Test: 1 vs. 2	0.025	0.138	0.651	0.013	0.018	0.079	**<0.01**	**<0.01**
3. Independent model (3 parameters)	-366.000	-328.256	-360.794	-360.161	-376.681	-349.693	-376.055	-357.197
i. Test: 2 vs. 3	**<0.01**	**<0.01**	**<0.01**	**<0.01**	0.757	**<0.01**	0.185	**<0.01**
Model of best fit:	2	2	2	2	3	2	1	1

Note: numbers in bold indicate p values that were significant at the Bonferroni-corrected alpha level, .006

Table 2.3: Related to Figure 2.3b: Perceived milkiness task with real stimuli - log likelihood values for each participant, with nested hypothesis test p-values comparing the saturated model with the additive and the additive model with the independent.

	Observer							
Model/comparative test	BC	CE	FN	LK	LS	NG	PB	PCM
1. Saturated model (15 parameters)	-222.145	-190.592	-148.960	-207.479	-189.531	-162.543	-218.127	-153.237
2. Additive model (6 parameters)	-224.996	-199.817	-161.446	-213.783	-194.265	-168.264	-225.167	-160.652
i. Test: 1 vs. 2	0.769	0.030	<**0.01**	0.181	0.395	0.247	0.120	0.096
3. Independent model (3 parameters)	-373.299	-375.616	-367.284	-367.522	-374.910	-372.053	-376.494	-374.249
i. Test: 2 vs. 3	<**0.01**	0.072	<**0.01**	<**0.01**	0.053	<**0.01**	0.056	<**0.01**
Model of best fit:	2	3	1	2	3	2	3	2

Note: numbers in bold indicate p values that were significant at the Bonferroni-corrected alpha level, .006

2.3.7 Experiment 2: Rendered stimuli MLCM results

Table 2.4: Related to Figure 2.4a: Perceived tea strength task with rendered stimuli - log likelihood values for each participant, with nested hypothesis test p-values comparing the saturated model with the additive and the additive model with the independent.

	Observer				
Model/comparative test	AC	EL	PM	YB	EG
1. Saturated model (15 parameters)	-177.512	-243.634	-274.148	-207.100	-278.612
2. Additive model (6 parameters)	-203.431	-252.788	-294.740	-238.793	-291.458
i. Test: 1 vs. 2	<**0.01**	0.032	<**0.01**	<**0.01**	<**0.01**
3. Independent model (3 parameters)	-248.444	-283.030	-370.814	-339.535	-302.867
i. Test: 2 vs. 3	<**0.01**	<**0.01**	<**0.01**	<**0.01**	<**0.01**
Model of best fit:	1	2	1	1	1

Note: numbers in bold indicate p values significant at the Bonferroni-corrected alpha level, .01

Table 2.5: Related to Figure 2.4b: Perceived milkiness task with rendered stimuli - log likelihood values for each participant, with nested hypothesis test p-values comparing the saturated model with the additive and the additive model with the independent.

	Observer				
Model/comparative test	AC	EL	PM	YB	EG
1. Saturated model (15 parameters)	-186.298	-250.521	-163.848	-239.655	-234.902
2. Additive model (6 parameters)	-199.496	-260.943	-185.274	-252.568	245.066
i. Test: 1 vs. 2	<**0.01**	0.013	<**0.01**	<**0.01**	0.016
3. Independent model (3 parameters)	-201.976	-267.885	-186.579	-260.269	-304.750
i. Test: 2 vs. 3	0.175	<**0.01**	0.456	<**0.01**	<**0.01**
ii. Test: 1 vs. 3	<**0.01**		<**0.01**		
Model of best fit:	1	2	1	1	2

Note: numbers in bold indicate p values significant at the Bonferroni-corrected alpha level, .01

Preface to Chapter 3

In the last chapter, we investigated how the manipulation of light scattering and absorption within volumes was perceived by observers. We used two separate experiments to investigate whether the visual system makes approximations of a simplified model of light scatter and absorption, or whether the use of pseudocues was more likely. The conclusions reached tend to corroborate the current consensus as to how the visual system interprets physical gloss (Chadwick & Kentridge, 2015; Fleming, 2014). However, not all translucent materials are found in substantial volumes. Many materials have a layer of translucent material at the surface, such as skin, vegetables, cheese, marble, plant material such as glossy leaves, and plastics. This change in composition of the material structure could potentially affect the way in which these materials are perceived or interpreted by the visual system. Therefore in the next chapter we replicate the study of the previous chapter where both light scattering and absorption are manipulated, but with stimuli involving translucent layers.

Chapter 3

Perceptual un-mixing in layers of translucent coating

Translucence is a defining characteristic of many natural materials. In a previous study, we investigated how perceived translucence of volumes of liquid related to its physical components. Many natural materials incorporate partially translucent layers near their surface. Such materials are also often glossy; indeed, the underlying driving determinants of gloss and translucence are physically similar as they both depend on the degree of light scattering, whether from or beneath a surface. In this study, we investigate how properties of such glossy materials are perceived when gloss is produced in a layer of coating. Observers completed a maximum likelihood conjoint measurement task, where on each trial two images of metal pans covered in a scattering and absorbing layer were presented, each varying in the amount of light the layer absorbed and scattered. Observers were asked to complete the task twice - once, making judgements of darkness regardless of cloudiness, and a second time making judgements of cloudiness regardless of darkness. Observers' perceptual estimates of the layers of coating relied on a complex combination of both of the physical variables. Judgements of darkness were not independent of cloudiness, and vice versa, suggesting that perceptual constancy of the estimated

parameters of the physical variables is affected.

3.1 Introduction

In the real world, objects and surfaces we encounter in everyday life are often made of layers of different materials, which are neither completely opaque nor transparent. When we perceive shiny surfaces, visual changes in these surfaces are usually a result of changes in the upper layers of the material rather than of the underlying material. Fruit flesh, marble, granite, uncooked meat and waxy leaves are all examples of shiny materials where the layers of material at the surface - often partially translucent - are responsible for perceived shine, and where changes at the surface result in changes in perceived shine.

In this experiment, we wanted to investigate whether using an approach taken to analyse bulk liquid volume materials (Chapter 2) - the perception of which likely involves the use of pseudocues (Chadwick & Kentridge, 2015) - might also apply when looking at layers of coating in relation to perceived gloss. How do our perceptual judgements of surfaces vary as a result of changes in these layers at the surface of a material?

We vary physical gloss by manipulating the scattering and absorption of light within a layer of coating. A set of metal pans with a solid layer of translucent coating was rendered in computer graphics, varying in two parameters: the levels of absorption and scatter of the layer of coating. A two-alternative forced choice method and maximum likelihood difference scaling analysis (MLDS; Knoblauch and Maloney 2012) was used to find an approximately linear perceptual scale of each physical variable. A group of observers were asked to complete one of two tasks - again a two-alternative forced choice design, making decisions about either what they might perceive as the 'dustiness or cloudiness' or 'darkness' of a pair of pots, while both physical parameters were varied. Maximum likelihood conjoint measurement analysis (Knoblauch & Maloney, 2012) was applied to determine the contribution of the physical variables to perceptual judgements. We used the terms 'dustiness' and 'cloudiness' to act as the antithesis to 'glossy' - as the more diffusely light is scattered from a surface, the less glossy it looks, and in this instance, the more cloudy/dusty the objects appeared. In contrast, increasing absorption (or darkness) of a surface has been shown to increase perceived gloss (Beck, 1964; Harrison & Poulter, 1951).

By investigating the effects of these parameters on perceptual judgements of layers of coatings, we can also explore potential implications for the nature of our perception of translucence.

3.2 Method and Results

3.2.1 Stimuli

Simulated images of physically realistic metal pots were created in Blender v2.68, an open source 3D computer graphics program which can model 3D scenes and objects. Each pan was coated in a light-scattering and absorbing layer of uniform thickness, within which light scattering and absorption could be manipulated in a physically realistic way. The scene was illuminated using a real world lighting probe based in a kitchen, to produce realistic reflections in the mirrored surface of the pots. Images were rendered using LuxRender, a raytracing program which provides a way of generating three-dimensional (in appearance) graphics. LuxRender is based on PBRT (physically based rendering software), and simulates the propagation of light through the scene according to accurate physical equations. It simulates physical properties of materials, including their reflective properties and, importantly, transport within volumes of materials. The images produced are of photographic quality as they retain the range of intensity of the original scene at arbitrarily high spatial resolutions.

The original stimulus set contained 165 rendered images. The properties of the coating were varied by manipulating the degree of physical light absorption (changing the apparent colour of the pan from a light silver to a dark gunmetal) and physical light scattering (such that the coating appeared dirtier with a higher degree of scatter). The numerical values of the parameters defined corresponded directly to the material property simulation - values of scatter specify the probability of a light scatter event per meter of travel, and values of absorption define the proportion of light absorbed per metre of travel (although the units themselves are arbitrary). Values for scatter and absorption were much higher for this set of stimuli than those in Chapter 2, as the smaller dimensions of the layer of coating compared to the volumes of tea meant that similar levels of scatter and absorption were not visible. There were 11 levels of physical scattering and 15 levels of physical absorption.

For MLDS testing, the set was reduced to 11 levels of each variable, with large step sizes to encompass the full range of stimuli generated. Following MLDS analysis, the set was reduced to 10 levels of physical absorption and 10 levels of physical scattering. These levels were both in the range of 0-72 units, where the unit is an arbitrary index of refraction per meter travelled, in increments of 8. There was therefore a total of 100 images in the experimental stimulus set. Extensive piloting then determined the optimum task difficulty for MLCM within this range, by setting the base level and step sizes for each variable. The base level of the scattering parameter was 0, and the base level of the absorption parameter was 16. A step size of 2 was used for scatter, and 1 for absorption, with 4 steps made overall: this produced a total of 4 levels used for each parameter within the stimulus set (see Figure 3.1).

3.2.2 Apparatus

Experimental software was written in Matlab. Stimuli were presented on a gamma-corrected ViewSonic 17 display monitor (1064x768 pixels, with refresh rate of 100Hz), controlled by means of a CRS Visage (Cambridge Research Systems). Responses were made via a standard keyboard. Observers were seated approximately 50cm away from the screen in a blacked out cubicle, and were not required to use a chin rest, but asked to sit at a comfortable distance from the screen whilst maintaining viewing distance.

3.2.3 Statistical software

All calculations were performed using Knoblauch and Maloney's (2012) MLDS and MLCM program packages for R (v2.11.1).

3.2.4 Observers

Four observers participated in the MLDS experiment, and five different observers participated in the MLCM experiment - all were aged 18-25, and had normal or corrected-to-normal vision.

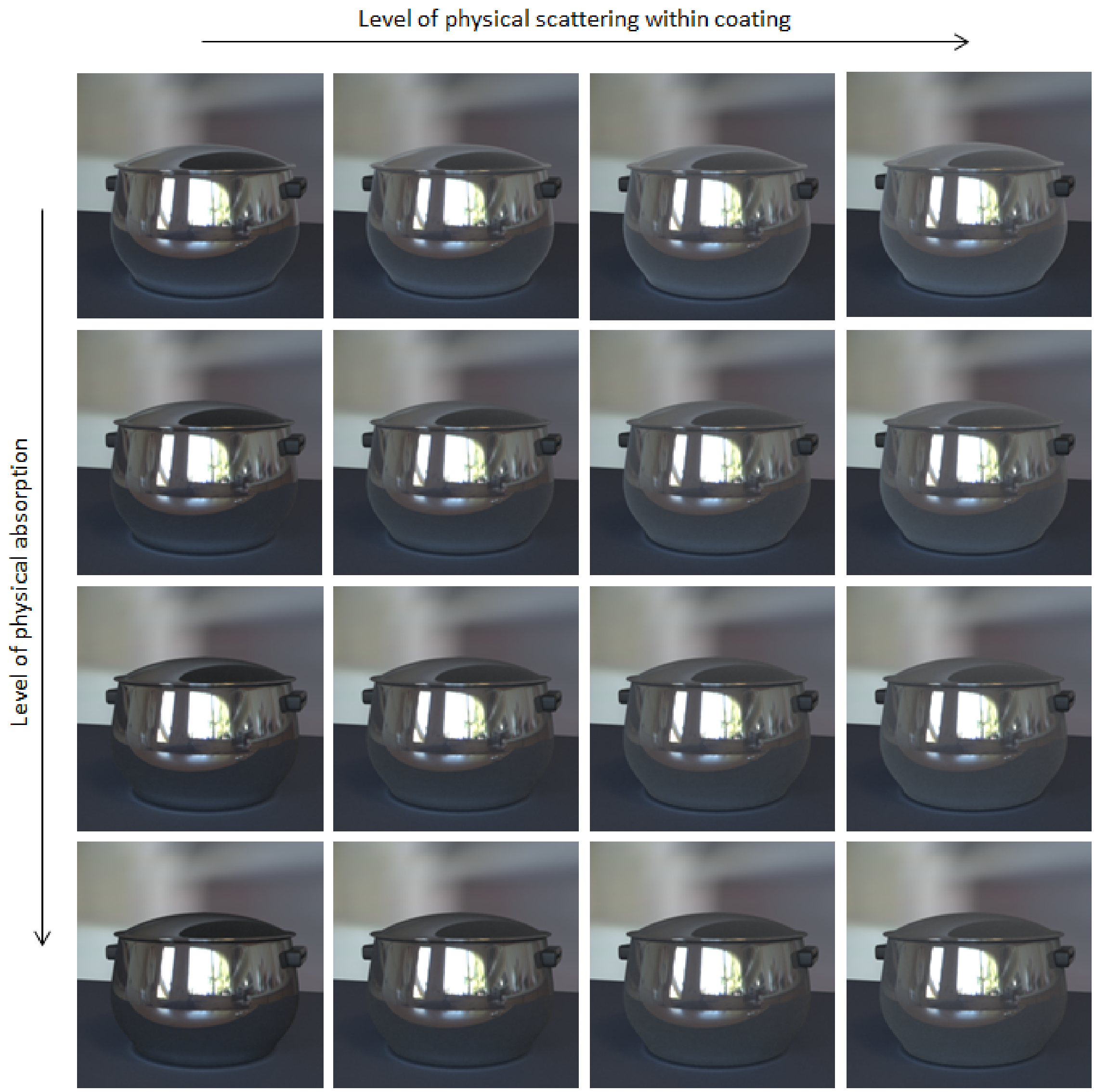

Figure 3.1: The full stimulus set used. Physical scatter varies along the x axis, and physical absorption along the y axis.

3.2.5 Maximum likelihood difference scaling method

Maximum likelihood difference scaling measures how perceived material properties vary as a function of a physical scale, and estimates a function relating the physical scale to perceptual parameter estimates. Perceptual differences for stimuli varying in one physical dimension are estimated, and a scale demonstrating how changes in the physical dimension translate to perceived differences is generated. Two pairs of stimuli are presented simultaneously in each trial (known as quadruples). Observers are asked to compare the two pairs, and make a judgement about which pair shows a greater perceptual difference. Only one parameter is varied within a single experiment, and the other is kept constant at a low level to avoid interference. The task is performed twice, once for each physical parameter used. Quadruples were chosen in accordance with the methods of Knoblauch and Maloney (2012) - only non-overlapping quadruples were used (that is, for the four values, a, b, c, d, of the parameter being manipulated, a<b<c<d) - and the maximum likelihood perceptual scales were then computed using the Knoblauch and Maloney MLDS package for R.

3.2.6 MLDS procedure

All observers completed both the scattering and absorption conditions of the MLDS task on consecutive days to avoid fatigue. Completion of the two conditions by each observer was counterbalanced. On each trial, observers were asked to judge - depending on the condition - which of the two pairs had a greater difference in either the 'darkness' of the pot or what they might perceive as the 'dustiness' or 'cloudiness' of the pot. The level of the irrelevant physical parameter was kept constant: scatter at 8, and absorption at 32, to minimise potential interference. Both dimensions had ranges of 0-80. 330 combinations were presented in a randomised order, and repeated for 3 blocks, giving 990 trials in total. The ordering of each pair - left to right, which had the greatest difference - was also randomised.

3.2.7 MLDS results

Figure 3.2 shows parameter estimates for perceived darkness and dustiness/cloudiness. After additional piloting to determine the optimal task difficulty, the levels of the parameters used in MLCM (in relation to those in the graphs above) were 4, 5, 6 and 7 for scatter,

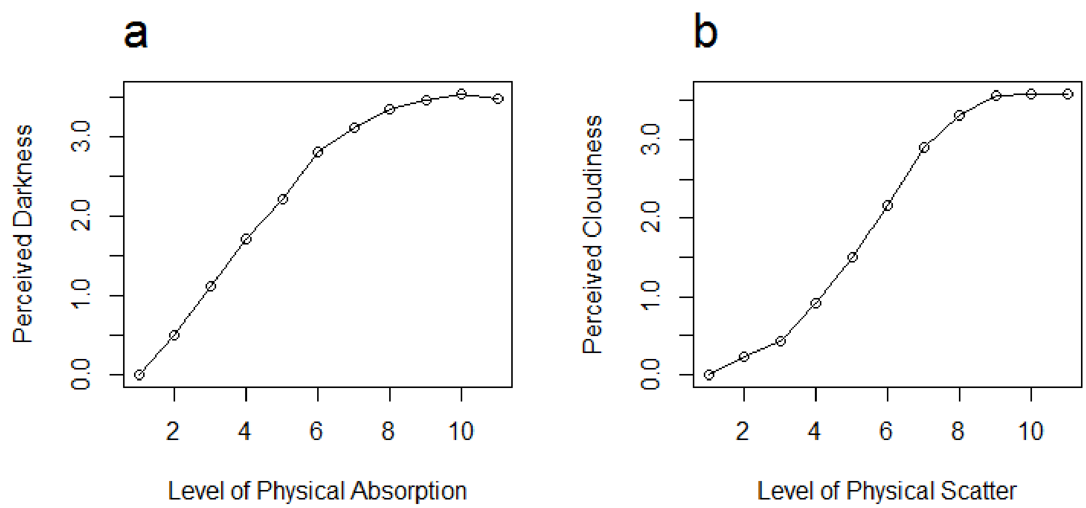

Figure 3.2: a) Perceived darkness (in d', on the y axis) as a function of physical absorption (x axis), normalised and averaged across all participants, and scaled to the average highest value. b) Perceived cloudiness (d', on the y axis) as a function of physical scatter (x axis), normalised and averaged across all participants, scaled to the average highest value.

and 2, 3, 4 and 5 for absorption.

3.2.8 Maximum likelihood conjoint measurement models

Maximum likelihood conjoint measurement aims to determine whether two different physical parameters contribute to a single perceptual dimension of judgement. The technique involves presenting pairs of stimuli to observers while manipulating both physical parameters from trial to trial. Observers are instructed to make a judgement about only one perceptual dimension (in this case, 'darkness' or 'cloudiness/dustiness' of the pots), reporting which of the pair is 'darker' or 'more dusty/cloudy'. This task can then be repeated with the same set of stimuli, with instructions to make perceptual judgements about the other dimension. The patterns of response can then be used to calculate how observers' perceptual judgements vary as a function of the physical parameters (for instance, whether perceived cloudiness varies as a function of physical absorption as well as physical scattering). The model then predicts the individual decisions that participants would make about a wide range of stimuli, compares this to the actual data obtained, and finds the parameters of the model that produce the best fit.

Maximum likelihood conjoint measurement provides three different ways of fitting the data: independent, additive, and saturated models. An independent model acts as an observer able to completely disentangle the effects of the two physical parameters might: perceptual judgements of parameters are assumed to be completely independent of one another; the model fits three hypothetical parameters (one for each level of the parameters used). An additive model assumes that an increase in the second 'irrelevant' parameter produces a simple confound effect, in that it would simply transpose the judgements obtained. That is, an increase in absorption might increase the perceived cloudiness, but by a consistent and predictable amount. The model therefore fits six hypothetical parameters; one for each combination of the levels used. A saturated model makes no assumption of linearity, and allows for complex non-linear interactions to be taken into account. This model fits to fifteen hypothetical parameters - one for each different combination of the levels used. A log likelihood value is produced for each model.

3.2.9 MLCM procedure

All observers completed both the scattering and absorption conditions of the task on consecutive days to avoid fatigue, and completion of the two conditions by each observer was counterbalanced. Each condition consisted of 4 blocks of 240 trials (960 trials in total), where in an individual trial one pair of stimuli was presented on the display screen (with the pair of images shown side by side). On beginning a block, observers were presented with a black screen before a pair of stimuli was shown for a maximum of 6 seconds. This was followed by an inter-trial interval (a black screen). The observer made a judgement about whether the stimulus on the left or right was darker or dustier, regardless of the level of the other parameter, and indicated their response by pressing one of two keys on the keyboard. The program waited on the inter-trial interval screen for a response, and the next trial was initiated immediately upon receiving the response. Observers could indicate their judgement before the end of the stimuli presentation time if they did not require the full 6 seconds. The order of trials was randomised within each block, and the ordering of each pair (left to right) was also randomised.

3.2.10 MLCM results

The data were fitted to the three models, generating three log likelihood values for each participant. A nested hypothesis test was performed on each set of log likelihoods. The nested hypothesis test makes comparisons to see if a less parsimonious model fits the data significantly better than a more parsimonious model. Tables 3.1 and 3.2 show the models of best fit for all observers, and the saturated model for each observer is plotted in Figures 3.3 and 3.4. For perceived cloudiness, four of the five observers' results were best fit by a saturated model, and the remaining observer gave a best fit for an additive model. This suggests that observers' responses for perceived cloudiness were dependent on a complex combination of both the two variables. For perceived darkness, three of the five observers were best fit by a saturated model, and two were best fit by an additive model. This suggests that the estimation of darkness was dependent on both absorption and scattering of light, and for three observers the contributions of the two physical variables were complex while for the remaining two the contributions were more straightforward.

While observers could make judgements of each property, and thought the two distinct, the estimation of each was dependent to some extent on the 'irrelevant' physical parameter. For perceived darkness, the results weigh slightly more in favour of an additive model. Perceived cloudiness is clearly dependent on the two physical variables in a more complex manner - with estimates changing with variation in both physical scattering and physical absorption. Overall, as physical scatter increased, perceived darkness decreased, and as physical absorption increased, the level of perceived cloudiness decreased, with some fluctuations in the patterns of perceptual responses along the two dimensions.

Generally, increases in perceptual estimates corresponded to increases in the physical dimensions; however there were interactions of physical dimensions affecting perceptual estimates when the two physical dimensions were manipulated simultaneously. The differences between participants imply that different weightings of pseudocues were being employed - indeed, as can be seen in Figure 3.3, observer EG perceived an overall increase in cloudiness when physical absorption was increased. This suggests that this observer was making judgements based on completely different pseudocues, or employing a differently weighted internal model.

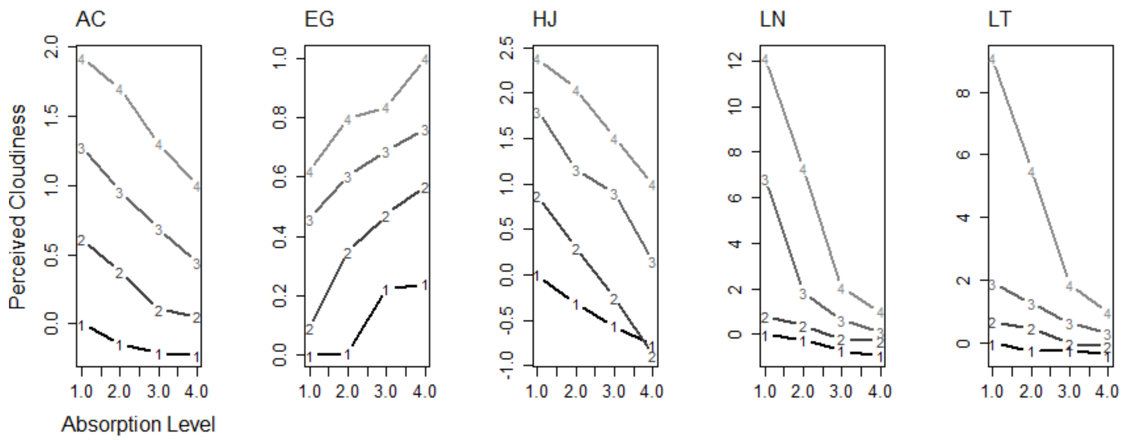

Figure 3.3: Perceived cloudiness of each individual (in d', on the y axis) as a function of physical absorption (x axis). Numbers 1-4 within the plots denote low to high levels of physical scatter.

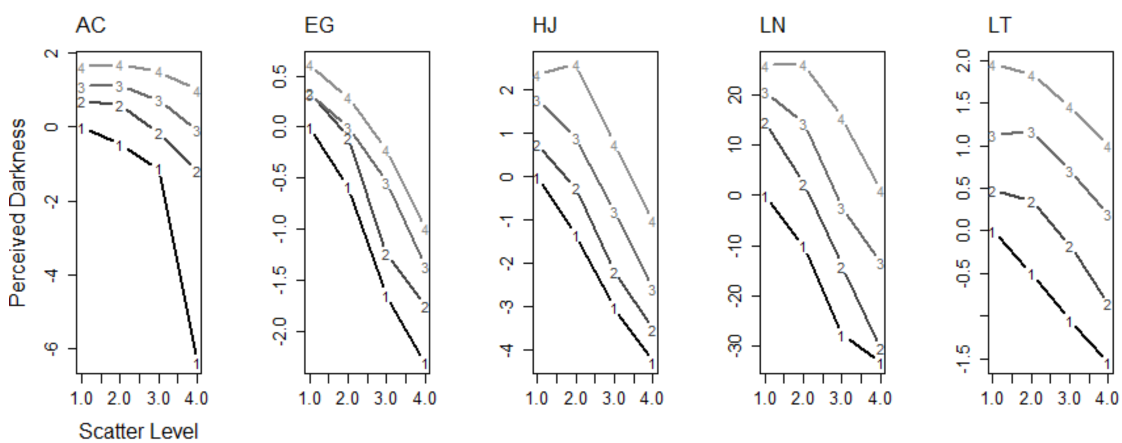

Figure 3.4: Perceived darkness for each individual (in d', on the y axis) as a function of physical scatter (x axis). Numbers 1-4 within the plots denote low to high levels of physical absorption.

Table 3.1: Perceived cloudiness task - log likelihood values shown for each participant, with nested hypothesis test p-values comparing the saturated model with the additive, and the additive model with the independent.

	Observer				
Model/comparative test	HJ	LT	EG	AC	LN
1. Saturated model (15 parameters)	-165.300	-203.089	-410.572	-229.922	-193.835
2. Additive model (6 parameters)	-183.047	-242.217	-415.708	-244.792	-226.894
i. Test: 1 vs. 2	**<0.01**	**<0.01**	0.329	**<0.01**	**<0.01**
3. Independent model (3 parameters)	-375.080	-357.534	-472.110	-328.615	-398.502
i. Test: 2 vs. 3	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**
Model of best fit:	1	1	2	1	1

Note: numbers in bold indicate p values that were significant at the Bonferroni-corrected alpha level, .01

Table 3.2: Perceived darkness task - log likelihood values shown for each participant, with nested hypothesis test p-values comparing the saturated model with the additive, and the additive model with the independent.

	Observer				
Model/comparative test	HJ	LT	EG	AC	LN
1. Saturated model (15 parameters)	-211.946	-369.150	-299.684	-265.568	-380.499
2. Additive model (6 parameters)	-220.280	-373.860	-314.166	-288.200	-389.525
i. Test: 1 vs. 2	0.054	0.399	**<0.01**	**<0.01**	0.035
3. Independent model (3 parameters)	-543.924	-451.263	-611.453	-422.575	-549.010
i. Test: 2 vs. 3	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**
Model of best fit:	2	2	1	1	1

Note: numbers in bold indicate p values that were significant at the Bonferroni-corrected alpha level, .01

3.3 Discussion

There was agreement across all observers in the overall effect observed - as physical scatter increases, perceived darkness decreased, and as physical absorption increased, perceived cloudiness decreased. This result is interesting as one might have assumed that as one physical parameter increased, it might also result in an increased perception of the other dimension - however it seems that, for example, as absorption increases, darkness is increasingly attributed to absorption, and estimates of scatter decrease, perhaps as overcompensation. It is also likely to reflect the finding that surface lightness significantly affects perceived gloss: manipulating the scattering of light within the surface coating layer effectively changes the physical gloss of the surface, and increasing absorption reduces surface lightness. As darkness of the surface increases, the surface is perceived to be shinier, even when physical gloss has been manipulated. The unusual result of participant EG may be due to application of an alternative pseudocue in making judgements of cloudiness, or different weightings within an internal model: perceived cloudiness increases with increased absorption, in contrast to other observers. It might well be that some element of contrast between the absorption and scatter of the layer induced this response, whereas all other observers may be attributing some of the darkening of the surface of the pot to a decrease in scattering (as the scattering parameter introduces a whitish appearance which may be interpreted as lightening of the surface).

The perceived cloudiness of the surface was best explained with a saturated model for four observers, and with an additive model for one observer. The perceived darkness of the surface was best explained with a saturated model for three observers, and two showed best fit to an additive model. This demonstrates that, in our percepts of darkness and cloudiness in a layer of coating, judgements are based in a complex way on the physical parameters. An independent model was not sufficient for any observer, showing that observers cannot be making estimates of the underlying physical parameters of scattering and absorption of light. For some observers, the complex interaction is not predictable or separable, whereas in others there is a simpler additive or transformative effect. This implied that we are not making estimates of darkness or cloudiness on the basis of either absorption or scattering of light alone; instead, there is some effect of each to the other when making perceptual judgements. We cannot, or do not, compensate fully for any perceptual effects generated by the supposedly irrelevant parameter. This suggests the use of more general pseudocues, possibly driven by complex spatially constrained contrast image statistics, obtained by parsing the image. Differences between observers indicate that observers weight pseudocues differently depending on the judgement being made, or use different pseudocues when dealing with ambiguous scenes.

These results support the conclusion that observers are not making estimates of the underlying physical light transport properties of the materials, and can also provide insight into the way in which visual judgements are made of these solid materials. The overall patterns of judgements of the layered stimuli found that slightly more observers showed best fit to an additive model compared to the findings obtained with rendered volumes

(Chapter 2); however, it is not clear that any conclusions can be drawn from this, as the differences are minimal. These stimuli are of course very different, in that reflections from the surrounding environment are visible, and more information is available from the scene via reflections in the layers which could allow observers to more effectively assess the distinctness-of-image gloss or reflections. There may well be a great deal more information available to observers in making visual judgements of these thinner layers, enabling observers to better maintain constancy. However, the rendered tea stimuli were not completely physically accurate in terms of the material they were intending to simulate, so we cannot necessarily compare the two. Observers were very good at separating contributions of the physical variables with real volumetric scattering and absorbing stimuli (Chapter 2). It may simply be that observers are more practiced at the real scenario, as it is more familiar - therefore the disentangling, in spite of more complex physical interactions, is more easily achieved; however, it is not possible to draw any conclusion about whether observers are more or less able to separate contributions of physical variables in volumes or layers.

These results further illustrate the relationship between lightness and gloss. Surface lightness was already known to affect perceived gloss, but here we have shown that this effect is maintained when physical gloss is also manipulated, as perceived gloss increased as darkness increased even at lower levels of physical gloss (higher levels of physical scattering). For some observers, increasing darkness was sufficient to overcome the effect of physical scatter on perceived cloudiness, however for others it was insufficient. This implies the effect is relative, and again subjective, as observers may employ different pseudocues, or weight them differently.

Preface to Chapter 4

In the previous experiment, we concluded that translucent materials found in layers were generally perceived similarly to those found in volumes. Such materials are seen in changing environments in everyday life, and one of the most frequently occurring changes in the natural world is the direction of illumination. In the next chapter, we report two experiments investigating how perceived translucence of volumes and layers are affected when the most important physical determinant of translucence - the physical scattering of light within a material - is conjointly manipulated with the direction of illumination within the scene. The latter experiment touches on both perception of translucence and translucence of gloss, as varying scatter in a very thin layer is very similar to manipulating physical gloss (this is explored further in a later chapter). The algorithms used to render physical gloss and transparent layers are different, as when simulating layers of coating the object is genuinely rendered with a relatively thin volume transport layer above an opaque material, rather than with variations in the proportion of specular and diffusely reflected light from the surface. The resultant image may look slightly different were an algorithm manipulating surface gloss used, however this method of using a scattering layer applied to an opaque base material means that should we want to compare alternative materials, we can be sure of applying identical manipulations to differing base materials. Sensitivity to differences between images produced by varying these two different parameters has not been investigated by this thesis - it is possible that observers could detect differences at grazing angles, where a layer would become thicker with respect to viewing direction, however the layers used are very thin and we do not believe that these effects would be significant. In these experiments, we ask questions of 'dustiness' and 'cloudiness' as in these instances, those terms represent the opposite of shine, and we are psychologically asking questions of gloss rather than translucence as we do not ask observers to make judgements of the nature of a translucent layer. Indeed, observers are not aware that there is a translucent layer present when this method is used, as only the scattering element is visible - there is no incident reflectance from the surface of the layer of coating itself, all specular and diffuse reflectance comes from the base material or scattering element. The effect of absorption in a layer is identical to a change in absorption of a surface, and it would be possible to make identical resultant images using two different algorithms (which is less likely for manipulations of scatter using different algorithms). Again, the data are considered in terms of perception of reflectance rather than in terms of perception of layers as observers are not asked to make judgements of translucent coating layers.

Chapter 4

The perception of translucence: interactions of light direction and subsurface scatter

Many natural materials, including plant materials, skin, marble, milk and meat can be characterised to some extent as being volumes, or having subsurface layers, of translucent material. The physical basis of such translucence is the extent to which light is propagated and scattered as it travels through a material. Human observers are sensitive to changes in translucence, and have some degree of constancy for it as viewing contexts change. However, it has been shown that changes in lighting direction significantly alters the perception of translucent volumes. In this study we set out to replicate this result and to investigate the nature of this interaction between translucence and illumination by employing the recently developed method of maximum likelihood conjoint measurement. In addition, and going beyond previous work, we tested this interaction both in volumes of translucent material and in objects with thin subsurface layers of light scattering material. In the first study, a group of observers completed a maximum likelihood conjoint measurement task where they were asked to make judgements of the cloudiness or translucence of objects, regardless of lighting direction, while physical scattering within the objects' volumes was manipulated alongside illumination direction. In the second study a group of observers was asked to make the same judgements, but now of solid objects with a subsurface scattering layer. We found that observers' judgements of translucence were inextricable from light direction in volumes of scattering liquid, but that judgements showed a greater degree of constancy under different illuminations when making judgements of scattering layers. However, light direction still contributed significantly to perceived translucence. We also demonstrate that altering the physical scattering of a layer of translucent coating material can be used to manipulate perceived gloss, as the main underlying physical determinant of gloss is essentially the same as for translucence - the extent of scattering of light from a surface.

4.1 Introduction

The human visual system is capable of distinguishing and identifying many different types of material and their properties, even within fractions of a second (Sharan et al., 2009). Translucence is one of these material properties, and study of the perception of translucence is becoming an expanding avenue of research in visual perception. Many objects are not completely opaque, and have a layer of translucent material or a coating which allows for some level of subsurface transport of light. The characteristics of light travelling through partially translucent volumes or layers can be defined using a bidirectional subsurface reflectance distribution function (BSSRDF).

We do have some degree of constancy in the perception of translucence. We are able to identify translucent materials under different illuminations and in different environmental contexts. Gkioulekas et al. (2013) found that multiple scattering of light contributes critically to the characteristic translucent appearance of food, liquids, and skin. The scattering of light within a volume (and therefore its translucence) is clearly very important for visual constancy and in identification of everyday materials and objects. However, Fleming and Bülthoff (2005) found that there can be large failures in constancy of the perception of translucence when lighting direction is varied. Observers perceive translucent objects as more opaque when front-lit, and more translucent when back-lit. This seems to be because, when front-lit, the only light that reaches the observer has entered and been reflected back from the translucent volume or surface layer, and has therefore been subjected to a higher degree of scatter than light which (when the object is back-lit) is transported through the volume or layer only once - or is at least scattered to a lesser degree before reaching the observer. Fleming and Bülthoff also explored the potential cues responsible for informing observers about whether a material is translucent or not. The physics of the transport of light through a volume is far too complex to estimate, and the information required to do so is not available to observers. Fleming and Bülthoff proposed that observers parse the scene into key regions, and gather image statistics from those regions.

In addition, Xiao et al. (2014) explored the interaction of shape, illumination and translucence constancy, also manipulating translucence appearance with lighting direction. They found that observers showed significant failures of translucence constancy across changes in lighting direction, and that this effect depended on two factors: the complexity of the three-dimensional shape of the object, and the translucency phase function - that is, the ratio of forward and backward scattering in a medium (as shown by Gkioulekas et al. 2013).

While the effect of light direction on perceived translucence has therefore been well established, we wanted to further investigate the precise nature of the interaction of the scattering of light and lighting direction in the perception of translucence. In the first section of this two-part experiment, we aimed to replicate the finding that translucence perception depends on lighting direction, in the presence of more explicit environmental cues to lighting direction. Furthermore, by using our combined method of maximum likelihood difference scaling and maximum likelihood conjoint measurement (MLDS and MLCM - Knoblauch and Maloney 2012, and Chapter 2), we aimed to characterise the precise nature of the interaction between lighting direction and the scattering properties of the material in the perceptual experience.

However, many objects encountered in daily life are not completely opaque or transparent, but have a layer of translucent material at the surface or a translucent coating (e.g. fruit flesh, skin, plastics, and ceramics). Therefore, in part two we were interested in exploring the effect of lighting and scatter when the translucent scattering substance was confined to a layer of coating on the surface of an object.

While the investigation of scattering volumes and lighting direction essentially explores the perception of translucency, investigating the interaction with layers of coating begins to address a slightly different question - that of perceived gloss. Manipulation of the scattering layer would also essentially equate to performing a manipulation of physical gloss. Translucence and gloss often co-occur: as Fleming and Bülthoff noted, translucent objects are often glossy, due to either the nature of the material itself (e.g. glass), or because of a layer of translucent glossy coating (e.g. metallic paints, ceramics, or plastics). As the tasks for the two experiments used different stimuli, the questions were framed differently to make them more appropriate for judgements made by observers - this means that they may touch on slightly different aspects of material perception by necessity. For part one, observers were shown pairs of glasses of milky liquid, and asked to make judgements of what they might perceive to be 'milkiness' or 'cloudiness'. For the layers of coating, observers were shown pairs of shiny pyrex dishes covered in a layer of scattering coating, and asked to make judgements about what they might perceive to be the 'cloudiness' or 'dustiness' of the surface.

4.2 Method

4.2.1 Apparatus

Experimental software was written in Matlab. Stimuli were presented on a gamma-corrected ViewSonic 17 display monitor (1064x768 pixels, with a refresh rate of 100Hz), controlled by means of a CRS Visage (Cambridge Research Systems). Responses were made via a standard keyboard. Observers were seated approximately 50cm away from the screen in a blacked out cubicle, and were not required to use a chin rest, but asked to sit at a comfortable distance from the screen whilst maintaining viewing distance.

4.2.2 Statistical software

All calculations were performed using Knoblauch and Maloney's (2012) MLDS and MLCM program packages for R.

4.3 Part 1: Light direction and scattering in volumes

4.3.1 Stimuli

Blender v2.68, an open source 3D computer graphics program which can model 3D scenes and objects, was used to simulate a glass tumbler containing a milky liquid sitting on a work surface amongst a number of cubes. These cubes, when seen with shadows, would provide additional cues to light direction in the scene. Images were then rendered in LuxRender, a raytracing program which provides a way of generating three-dimensional (in appearance) graphics. LuxRender is based on PBRT (physically based rendering software), and simulates the propagation of light through the scene in a physically realistic way (according to physical equations of the materials in the scene). Images produced using LuxRender are realistic and of photographic quality.

The original stimulus set contained 100 rendered images of the glass. The properties of the liquid were varied by manipulating the degree of physical light scattering, such that at higher levels of scatter the liquid appeared more milky and opaque. A real-world lighting image probe capturing a natural light distribution illuminated the scene, and two additional lamps (one to the rear of the glass, and one located behind the camera) were used to alter the overall appearance of the direction of illumination. The direction of lighting was manipulated by weighting the brightness of the two lamps, gradually exchanging brightness of one for brightness of the other. There were 10 levels of lighting direction (going from back-lit to front-lit), and 10 levels of scattering within the volume. MLDS was used to ensure that increases in scatter corresponded to equal differences in the perception of scatter. For the MLCM two-alternative forced choice task, the possible pairs were chosen on the basis of comprehensive piloting to determine the optimum task difficulty by setting the number of steps that each variable could make between the two stimuli in a pair at any point. The base level (-0.5 for lighting, 1.25 for scatter) and overall ranges within the set of stimuli were determined through the initial MLDS experiment. A step size of 1 was used for scatter, and 1 for light direction, and a total of 4 steps were used for each variable in the MLCM task (see Figure 4.1).

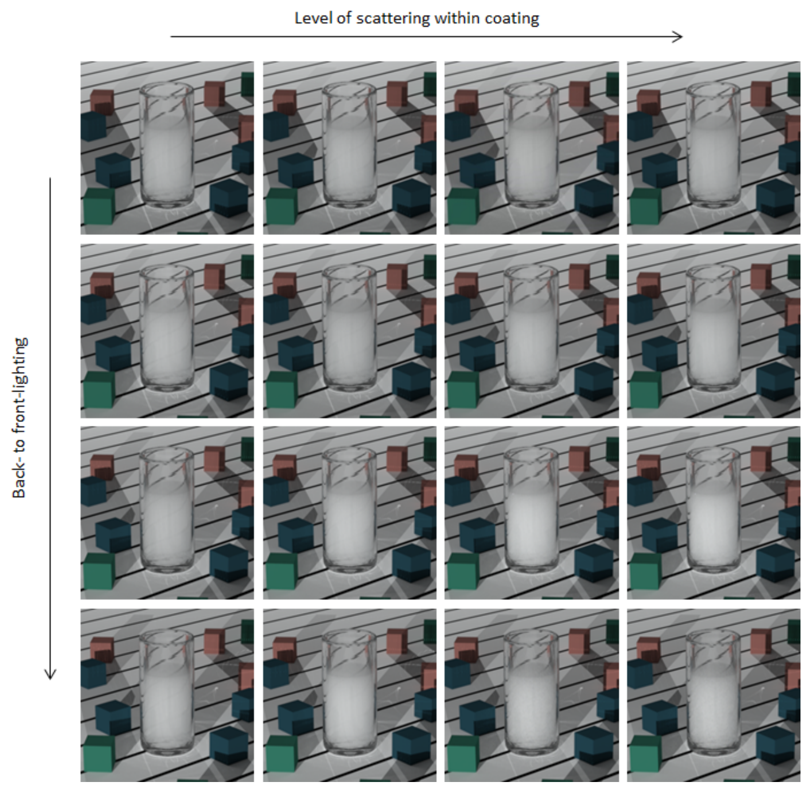

Figure 4.1: The full range of stimuli used in Experiment 1: light direction changes from back- to front-lit on the y-axis, and level of physical scattering varies on the x-axis.

4.3.2 Observers

4 observers participated in the MLDS experiment, and 5 in the MLCM experiment. All were aged 18-25, and had normal or corrected to normal vision.

4.3.3 MLDS procedure

The MLDS task involved presenting four images simultaneously, arranged in two pairs, varying in the level of scattering. Observers were asked to decide which of the two pairs showed a greater perceptual difference. The direction of lighting was kept constant within pairs at a neutral direction. A total of 210 combinations was presented in a randomised order, and repeated for 4 blocks (840 trials in total). The ordering of each pair - left to right, which had the greatest difference - was also randomised. The MLDS procedure was only performed for the scattering parameter as we were only interested in how perception of cloudiness changed, however the lighting direction dimension was checked to ensure that changes in lighting were detectable (and approximately equally discriminable).

4.3.4 MLDS results

Parameter estimates can be seen in Figure 4.2. Judgements increased approximately proportionally with increases in the physical dimensions, until a ceiling effect was reached with all participants. The higher end of the range of stimuli used was restricted for MLCM testing, to ensure that any potential effects were not a result of an inability to perceive differences in scatter. The range was reduced to levels 3-6, as seen in Figure 4.2. The number of lighting direction levels was also reduced to obtain an appropriate number of combinations.

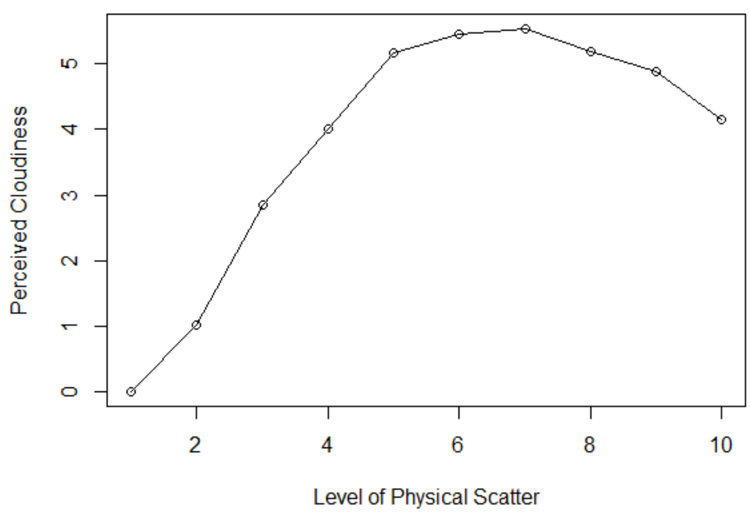

Figure 4.2: Perceived cloudiness (in d', on the y axis) as a function of physical scatter (x axis), normalised and averaged across all participants, and scaled to the average highest value.

4.3.5 MLCM procedure

The task consisted of 3 blocks of 240 trials where in an individual trial one pair of stimuli was presented on the display screen, the two images shown side by side. On commencing a trial, observers were presented with a black screen before a pair of stimuli was presented for a maximum of 6 seconds. This was followed by an inter-trial interval of a black screen. Observers made a judgement of whether the image on the left or right was cloudier/milkier regardless of the lighting direction. The program waited on the inter-trial interval for a response, and observers indicated their response by pressing one of two keys. Observers could make a judgement before the end of the 6 seconds given if they did not require the full time to inspect the images. The order of the trials was randomised within each block, and the ordering of each pair (left to right) was also randomised.

4.3.6 MLCM results

Three log likelihood values were produced for each participant; one for each of the three models being tested. A nested hypothesis test was performed on successive fits to determine the most parsimonious model for the data by calculating whether an increase in the number of parameters in the model explained a significantly higher proportion of the variance. A saturated model provided the best fit for the responses of all five observers (Table 4.1), demonstrating that judgements of cloudiness could not be separated from light direction. As seen in Figure 4.3, a great deal of consistency between participants is evident in the patterns of responses at all levels of the parameters. Perceived lighting direction patterns also showed remarkable consistency between participants, implying that observers were using the same cues or behaviourally indistinguishable models. Even in the presence of an explicit source of information for lighting direction, constancy of perceived light scattering could not be achieved by observers.

Table 4.1: Perceived cloudiness - log likelihood values shown for each participant, with nested hypothesis test p-values comparing the saturated model with the additive, and the additive model with the independent.

	Observer				
Model/comparative test	AC	SH	LN	SW	HH
1. Saturated model (15 parameters)	-139.284	-177.765	-152.496	-164.387	-215.624
2. Additive model (6 parameters)	-172.305	-215.431	-170.545	-187.741	-230.557
i. Test: 1 vs. 2	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**
3. Independent model (3 parameters)	-365.982	-350.057	-319.626	-308.209	-342.147
i. Test: 2 vs. 3	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**
Model of best fit:	1	1	1	1	1

Note: numbers in bold indicate p values that were significant at the Bonferroni-corrected alpha level, .01

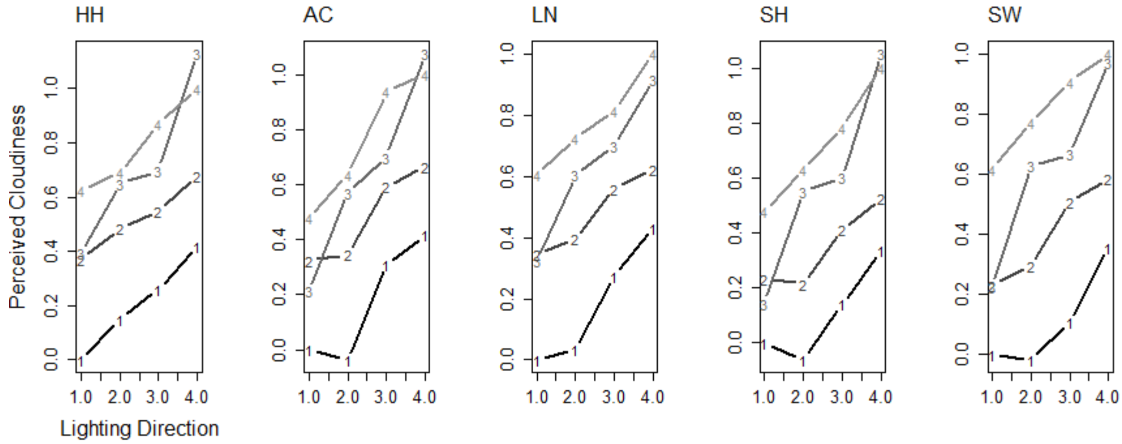

Figure 4.3: Perceived cloudiness for each individual (in d', on the y axis) as a function of lighting direction (x axis - numbers from 1-4 denote back- to front-lighting). Numbers from 1-4 within the plots denote low to high levels of physical scatter.

4.4 Part 2: Light direction and scatter in layers

4.4.1 Stimuli

The procedure in section 4.3 was used to create a set of simulated pyrex dishes with varying levels of scatter within the layer of coating, and varied lighting direction (from back- to front-lit). Pyrex was used because this is a common material in kitchenware, and LuxRender includes measured light transport parameters of the material. The original stimulus set generated comprised 143 rendered images of the pyrex dish, covered in a layer of coating. The properties of the coating were varied by manipulating the degree of physical light scattering, such that at higher levels of scatter the coating appeared dustier. There were 13 levels of lighting direction (going from back-lit to front-lit), and 11 levels of scatter (from 25-75 with step size 5). MLDS was used to ensure that increases in physical scattering within the layer corresponded to increases in perceived dustiness/cloudiness of the surface, and that changes in lighting from back to front were also perceived as changing from back to front. As stimuli were presented in pairs, the possible pairs were chosen on the basis of extensive pilot studies that determined the optimum task difficulty, by setting the number of steps that each variable could make between the two stimuli in a pair at any point. The base levels and overall ranges within the set of stimuli were determined through the initial MLDS experiment. A total of four steps were used for each variable in the MLCM task (see Figure 4.4).

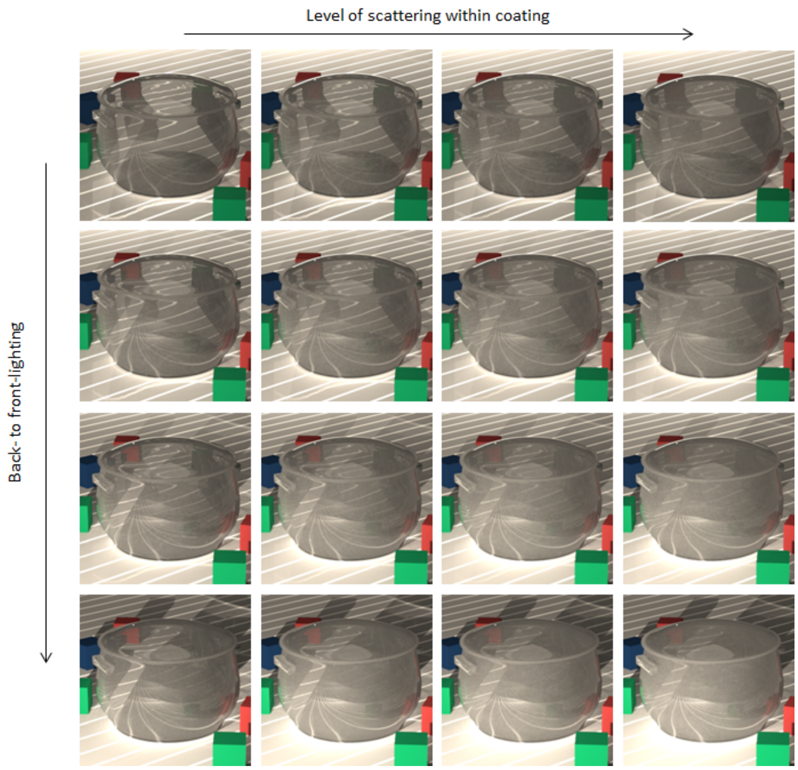

Figure 4.4: The full range of stimuli used in Experiment 2: light direction changes from back- to front-lit on the y-axis, and level of physical scattering varies on the x-axis.

4.4.2 Observers

4 observers participated in the MLDS experiment, and 9 in the MLCM experiment. All were aged 18-25, and had normal or corrected to normal vision.

4.4.3 MLDS procedure

The direction of lighting was kept constant. A total of 15 combinations was presented in a randomised order, and repeated for 40 blocks (600 trials in total). The number of blocks was increased to compensate for the smaller number of trials per block. The ordering of each pair - left to right, which had the greatest difference - was also randomised.

4.4.4 MLDS results

Parameter estimates can be seen in Figure 4.5. The full range of stimuli was used in the MLCM part of the experiment.

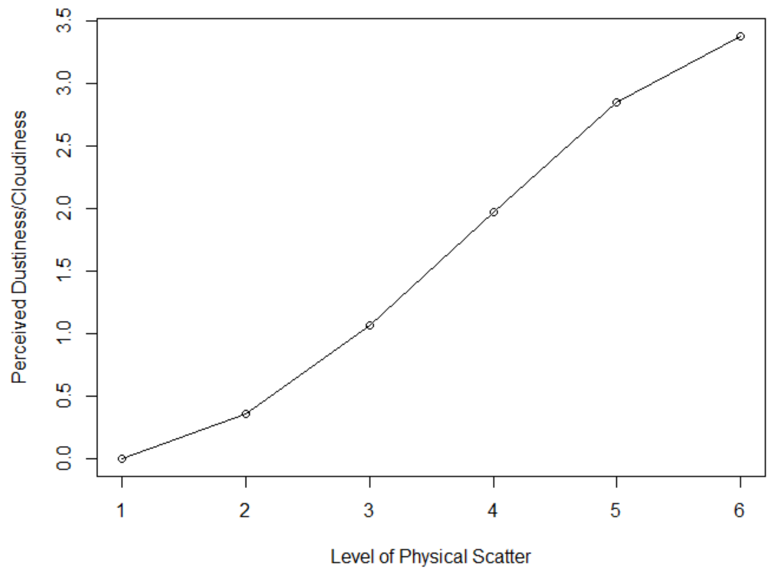

Figure 4.5: Perceived cloudiness (in d', on the y axis) as a function of physical scatter (x axis), normalised and averaged across all participants, and scaled to the average highest value.

4.4.5 MLCM procedure

The task consisted of 5 blocks of 240 trials, and the format of the experiment otherwise exactly replicated the MLCM procedure for section 4.3.

4.4.6 MLCM results

Three log likelihood values were calculated for each participant, and a nested hypothesis test was performed comparing the saturated and additive models, and the additive and independent models (Table 4.2). Patterns of responses between participants were again very consistent, although some observers seem to show a greater rate of increase in perceived dustiness at the lowest level of physical scatter, for the final step change in lighting direction (from front- to back-lit) which implies that these observers may be using similar sources of information to make judgements of scatter, that are slightly different to the sources being used by the other observers (see Figure 4.6). As opposed to the previous task, the best fit for all of the nine observers' data was the additive model. This suggests that observers could disentangle light direction and perceived dustiness/cloudiness, although could not do so entirely, but observers were better able to do so with layers compared with the results from the volumes experiment. A possible explanation for this is that scatter within layers of coating applied to a transparent material allows much more background or environmental information to be seen through the surface compared with larger volumes of scattering liquid, so more explicit information about the surrounding environment is available. Contours, surface, and edges of objects in the background of the scene are visible through the surface as well, so can better aid comparison of images when making judgements of dustiness/cloudiness; acting as an aid to compare, for instance, contrast image statistics.

Table 4.2: Perceived cloudiness - log likelihood values shown for each participant, with nested hypothesis test p-values comparing the saturated model with the additive, and the additive model with the independent.

	Observer								
Model/comparative test	AC	IAM	AMP	HH	LN	SCH	SH	SW	YB
1. Saturated model (15 parameters)	-169.369	-504.784	-82.373	-455.170	-168.966	-228.091	-422.718	-204.529	-513.966
2. Additive model (6 parameters)	-175.012	-513.596	-91.957	-458.276	-180.145	-233.283	-431.242	-211.801	-521.083
i. Test: 1 vs. 2	0.257	0.040	0.024	0.719	0.008	0.320	0.048	0.104	0.719
3. Independent model (3 parameters)	-744.241	-751.318	-749.095	-757.840	-733.226	-707.072	-725.760	-741.254	-763.471
i. Test: 2 vs. 3	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**
Model of best fit:	2	2	2	2	2	2	2	2	2

Note: numbers in bold indicate p values that were significant at the Bonferroni-corrected alpha level, .0056

4.5 Discussion

In this two-part experiment, we aimed to replicate the finding that lighting direction significantly affected observers' perceptual judgements of a transparent volume or surface, even with more explicit cues to the physical lighting direction present, and to further investigate the nature of this interacting relationship. Furthermore, we tested this with two different types of stimuli - volumes of liquid, and in layers of coating on a transparent object, in order to investigate potentially interesting effects with translucent volumes and surfaces varying in physical gloss.

An 'ideal' observer, according an inverse optics theory of vision, would have produced a pattern of judgements with a best fit to an independent model, as their perceptual judgements of one property would not have been affected by a second - however, in both parts of

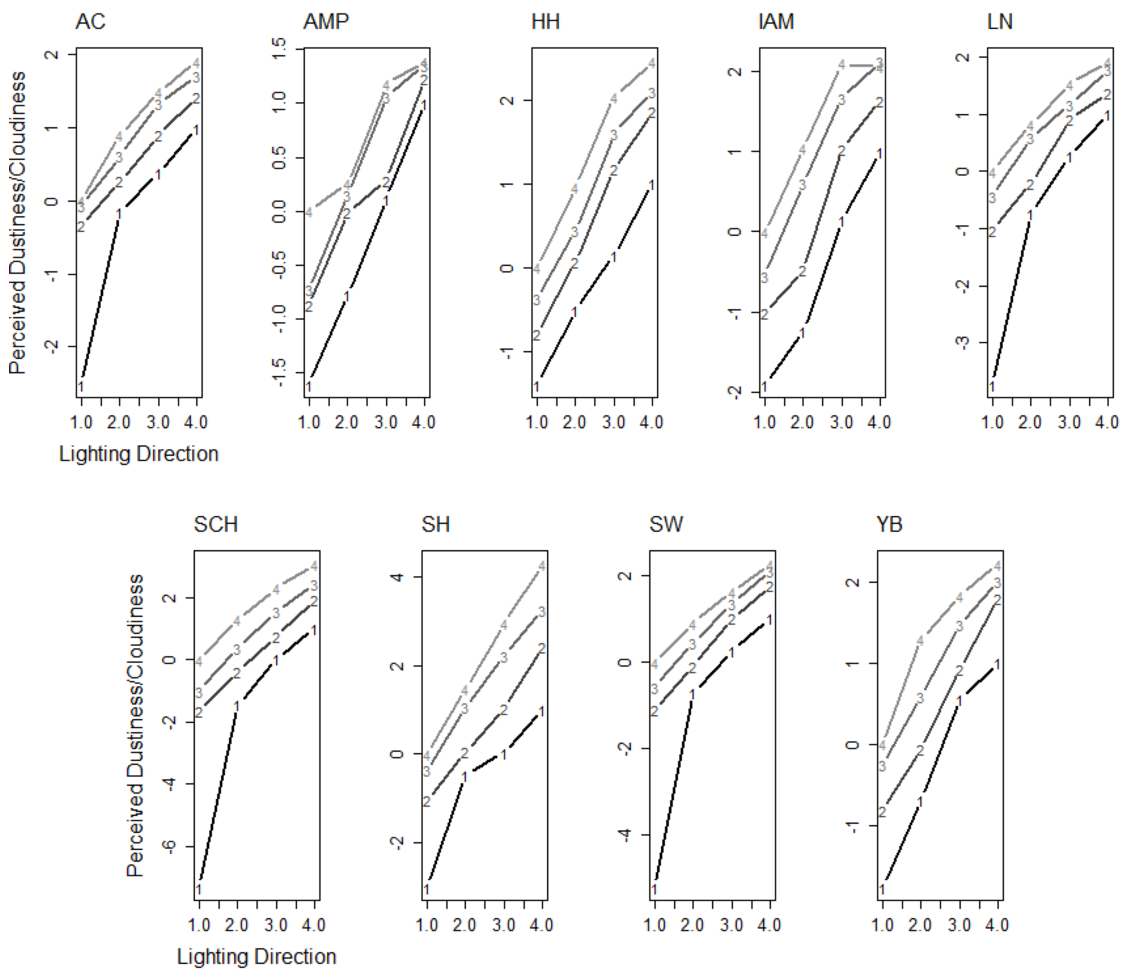

Figure 4.6: Perceived cloudiness (in d', on the y axis) as a function of light direction (x axis - where numbers 1-4 denote back- to front-lit). Numbers from 1-4 within the plots denote low to high levels of physical scatter.

this experiment this was found not to be the case. In fact, the pattern of responses found was comparable to earlier results looking at perceptual judgements of physical absorption and physical scatter. In the first half of this experiment, it was shown that observers demonstrated a best fit to a saturated model when making judgements of the scattering of light in a volume as the direction of lighting in the scene changed. That the relationship was not independent shows that the two physical dimensions manipulated in the experiment were both contributing to the observers' perceptual judgement; observers cannot be making veridical estimates of the underlying properties of light transport, as they have imperfect translucence constancy under changes of lighting direction. The extent of this entanglement of the physical dimensions in our perceptual judgement-making is exemplified by the fact that the pattern of responses showed a saturated pattern of best fit rather than an additive pattern - the two dimensions are confused to the extent that there is not a consistent way to map the physical dimensions onto the perceptual judgements. In the second part of the experiment, observers' judgements of light scattering in a layer of coating showed evidence of a best fit to an additive model. In this case, both of the physical dimensions were contributing to the perceptual judgements; however the 'irrelevant' dimension was contributing to the perceptual decision in a more predictable way across all levels of the lighting direction dimension. Therefore, participants were seemingly able to better distinguish the information available to them in the scenes. This might have been a result of the fact that the layer of scattering material is much thinner than the transparent volume, so background details were more visible through the layer, and these readily visible details may have offered more information in making comparisons between different images.

What should we expect of the relationship between scattering and perceived gloss? Surface roughness affects gloss as rough surfaces scatter more light than glossy surfaces. Therefore adding a scattering layer has the same effect on the underlying material, in decreasing the perceived gloss. However, in this case, it appears to be difficult to form firm conclusions about how manipulating the physical gloss and lighting direction affects perceived gloss. In the stimulus set, some front-lit objects show very clear reflections from the surface of the dish with few bright highlights, but some back-lit objects have bright - but small - highlights on the lid of the pot, despite not having as many clear reflections elsewhere. This suggests a more complex relationship between physical scatter, light direction, and perceived gloss. Increasing physical scatter appears to decrease perceived shine, however a change in lighting direction changes the types of gloss visible (Hunter 1937 - such as the distinctness-of-image in reflections, to brighter specular highlights and increased contrast gloss).

To try and address this question, a small number of observers (3) were asked to make some explicit comparative judgements of shine, while scattering and lighting direction were varied. The observers generally identified front-lit objects as looking shinier than back-lit, commenting that this produced more obvious reflections and specular highlights on the surface of the front-lit pot whereas back-lit pots had a very limited number of specular highlights mainly located at the very top edge of the pot lid. This agrees with previous findings: that larger, brighter highlights increases the area of the surface perceived to be shiny, that objects appear glossier with more visible highlights, and that when a highlight is some distance from the majority of the visible surface gloss ratings decrease (Beck and Prazdny 1981; Berzhanskaya et al. 2005; Fleming et al. 2003; Marlow & Anderson, 2013). Observers also generally perceived objects with higher levels of scatter - disregarding light direction - to be less shiny, which confirms the idea that manipulating scattering within the layer of coating can be equated to manipulating physical and perceived gloss. However, observers also commented that they sometimes found it difficult to interpret the variation in opacity. At lower translucence, they sometimes thought the object was shinier, as the reflected images were brighter. As only the scattering within the layer of coating was manipulated, distinctness-of-image gloss was constant between all stimuli, however changes in contrast gloss may have produced this confusion as well as changes in lightness. Furthermore, these stimuli were otherwise translucent, meaning that changes in scattering within the layer changed the opacity of the object itself as well as producing local surface effects; potentially resulting in the decision-making process to be a more cognitively difficult task. It therefore seems that in the context of these stimuli, the relationship between lighting and physical gloss becomes far more difficult to predict, particularly as the stimuli themselves were transparent.

A relationship between the physical dimensions and their perceptual correlates of the nature described above demonstrates evidence in favour of a theory of vision where observers might be taking in the 'gestalt' of the scene, and using a constellation of cues to inform their judgements. Observers might also be using pseudocues: learned methods of identifying particular properties of a surface which may involve comparison of a number of different sources of information, or estimates of the characteristics of the surface, or even judgements of the statistical patterns in the scenes. These might themselves be the way of implementing a holistic approach to perceptual judgements, since the pseudocues may incorporate changes in multiple aspects of the visual scene. Heuristics and pseudocues are, of course, less than perfect predictors. In some cases the pseudocue might be more informative and thus provide a more accurate judgement, but changes in the environment might confuse or alter the predictive power of these pseudocues. Such an approach also accounts for differences between observers. In terms of a gestalt, observers may be choosing to weight different sources of information in different ways, and prioritise one cue over another. Alternatively, the pseudocues adopted may be slightly different, so each observer's pattern of responses also differs - however, the level of consistency seen here between participants might imply that generally we use fairly similar pseudocues to make perceptual judgements. Indeed, we generally agree that a particular object appears to be shiny, but we may not always agree when the lighting or texture of the object is manipulated towards a 'grey area' of certainty, as we would be employing different criteria by which to judge the object.

Preface to Chapter 5

In the previous chapter, we addressed how one common contextual change of the environment, lighting direction, influences perceived translucence of volumes and layers. Besides potential influence of environmental changes, translucent layers are found on or within many different kinds of base material. In the next experiment, we use the same technique of simulating translucent layers, and apply these layers to a number of different base materials to investigate how observers' perceptions of these layers might be influenced by these base materials.

Chapter 5

Different base materials influence the perception of translucent coatings

Surface gloss is a characteristic property of many natural materials. Human observers are sensitive to changes in physical gloss. Layers of subsurface translucent scattering material have been shown to affect perceived gloss, as the determinants of surface gloss and subsurface translucence are very similar - both depend upon the extent to which a material scatters light. Perceptual constancy of gloss produced by such layers has been investigated under changes in illumination direction. The glossiness of many different types of objects are affected by translucent subsurface scattering layers. Here we ask whether there is something akin to perceptual constancy when considering the effects of identical scattering layers on different base materials. This study uses a technique developed in previous experiments simulating and manipulating a layer of scattering coating on perceived gloss. Here we investigate the differences in perceptual experience of translucent layers on different base materials, a transparent material (pyrex), a dielectric (stainless steel), and a ceramic. Three different groups of participants completed a maximum likeli-

hood conjoint measurement (MLCM) task. Analysis of the results allows the contributions multiple physical variables make to single dimensions of perceptual judgment to be estimated. Each group completed the MLCM task for one base material type, making judgements of darkness or cloudiness, while both the physical absorption and scattering of the subsurface layer were manipulated. The behavioural responses from the three conjoint measurement tasks enabled us to compare the estimates of the contribution of the two physical variables to percepts of the darkness and cloudiness of objects for the three different base materials. The results indicated that translucent base materials such as pyrex produced judgements from observers which were based in a more complex way on physical absorption and scattering. Judgements were also related to the physical variables in a complex fashion for the dielectric base material (stainless steel). We note that in both cases image cues to gloss, either reflected or transmitted, are available. The ceramic material produces almost no reflected image on its surface. Observers distinguished the contributions of physical absorption and scatter towards their perceptual estimates of darkness and cloudiness far more clearly on the ceramic base material. The implications for perception of translucence and gloss are discussed.

5.1 Introduction

When we perceive shiny surfaces, changes in perceived shine are caused by variation in a number of factors, including the physical scattering of light produced by the surface, surface darkness, and surface shape. Physical gloss is generally defined as the amount of specular reflectance from a surface, and is affected primarily by the amount of scattering of light - rougher surfaces scatter more light in multiple directions, and so look less shiny. Factors such as darkness and shape affect perceived gloss by altering characteristics of the specular highlights, such as the overall contrast between the highlights and the surface material and the size of the highlights (Chadwick & Kentridge, 2015). However, perceived gloss is not just altered by changes in physical scatter from the surface plane, but also in layers of coating or partially translucent material near the surface of the object. Light is reflected directly from the surface of the object, but also penetrates translucent layers of material and is reflected back indirectly to the observer, altering the amount of physical scattering produced by the surface, changing the perceived gloss of the object.

Such coatings and layers can be found on many different kinds of materials. Natural objects frequently have a layer of partially translucent material close to the surface, such as waxy shiny leaves. The human visual system is adept at identifying and distinguishing different materials using visual cues - we can easily tell the difference between wood and rock, and can distinguish real fruit or skin from visually very similar waxworks. We can even identify classes of materials correctly within fractions of a second (Sharan et al., 2009).

In this study, we wanted to investigate how perceived translucence in surface layers might be affected by applying the same surface layer to a number of different base materials. This is a novel method which has a particular advantage: by using a coating, we are able to assess the effects of generic scattering and absorption manipulations on disparate materials. It would not be possible to replicate this by manipulating values of scatter and absorption in each base material, as these values are not necessarily equivalent between the different materials; these manipulations need to be identical. We hypothesised that the same layer may affect perceptions of the glossiness of different materials in different ways because they may affect the pseudocues used in gloss perception (Chadwick & Kentridge, 2015) in different ways, and would thus produce a more complex model of perceptual judgements in relation to the physical parameters. Other materials might make visual cues more explicit and therefore aid perceptual decision making, and so produce less complex models of perceptual judgements as based on the physical parameters.

To investigate this question, we replicated a previous study of the effect of a layer of light-scattering material (Chapter 3) on perceived gloss, with two additional conditions where the base material was varied. These three familiar materials - ceramic, metal and pyrex - were rendered physically realistically with layers of coating varying in levels of absorption and scattering of light. An approximately linear perceptual scale within the set of stimuli was calculated for both physical parameters in each condition, by giving a group of observers a two-alternative forced-choice maximum likelihood difference scaling task (MLDS). Observers made decisions about which of two pairs of stimuli showed a perceptually greater difference in darkness or cloudiness. For each condition, a group of participants again performed a two-alternative forced-choice task making judgements of darkness or dustiness of the objects, and data were analysed using maximum likelihood conjoint measurement techniques (MLCM; Knoblauch and Maloney 2012). These techniques enable us to estimate the contributions of physical variables towards the perceptual experience by fitting the obtained data to three models, varying in the number of parameters estimated and therefore in the modelled complexity of interactions.

5.2 Method and Results

5.2.1 Stimuli

A pot shape was simulated using Blender, an open source 3D computer graphics program, and three pots of different base materials were created using LuxRender. LuxRender simulates physical properties of materials, including their light-transmitting and light-reflective properties. Each pot was coated in a layer with absorbing and scattering properties, and was set in a realistic scene. One set was created for each of the three base materials: pyrex, ceramic, and metal. Images were then rendered in LuxRender, also a raytracing program which provides a way of generating three-dimensional (in appearance) graphics.

The original stimuli sets generated comprised 100 rendered images of each base material, with the layer of coating varying in levels of absorption and scattering of light. The properties of the coating were varied by manipulating the degree of physical light absorption and physical light scattering. The manipulations were spectrally biased, such that at higher levels of absorption the coating appeared dark grey and at higher levels of scattering the coating appeared cloudy and white. A real-world lighting image probe employing natural spectral light distributions illuminated the scene. The parameter levels specified the proportion of scattering and absorption of the coating, in comparison with specified coating properties of high absorption or high scattering. The scatter and absorption units themselves define the probability of scatter through a metre of the substance, or the attenuation of light per metre of substance, respectively. The values of the parameters are therefore arbitrary. We set the probability of forward and backward scatter as being equal. The method of simulation was identical to that used in Chapter 3.

There were 10 levels of physical absorption and 10 levels of physical scattering within the initial stimuli sets. For MLDS testing, all levels of the variables were used. A perceptual scale for each parameter was derived, and the experimental stimuli sets were reduced to best reflect the range within the sets where observers perceived approximately equal increases. As stimuli were presented in pairs in the MLCM experiment, the optimum task difficulty was determined on the basis of comprehensive piloting by setting the base level for each variable, and the number of steps that each variable could make between the stimuli within a pair, so that perceptual differences were approximately equal. A total of 4 steps were made through the stimulus set for each parameter (see Figures 5.1, 5.2, and 5.3).

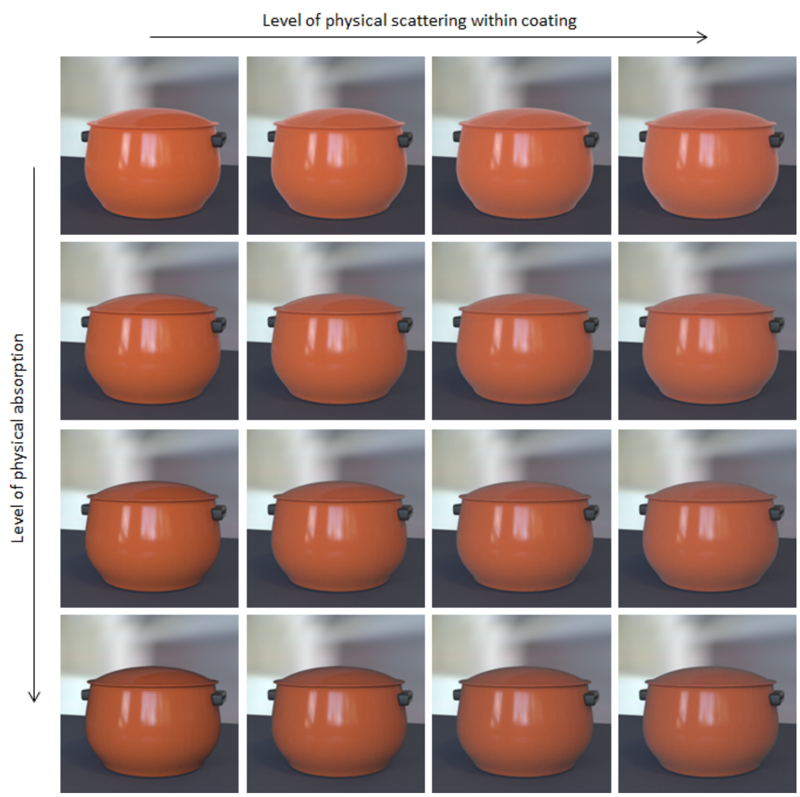

Figure 5.1: The full range of stimuli used; physical absorption varies along the y-axis, and level of physical scattering varies on the x-axis.

5.2.2 Apparatus

Stimuli were presented on a gamma-corrected ViewSonic 17 display monitor (1064x768 pixels, with refresh rate of 100Hz), controlled by means of a CRS Visage (Cambridge

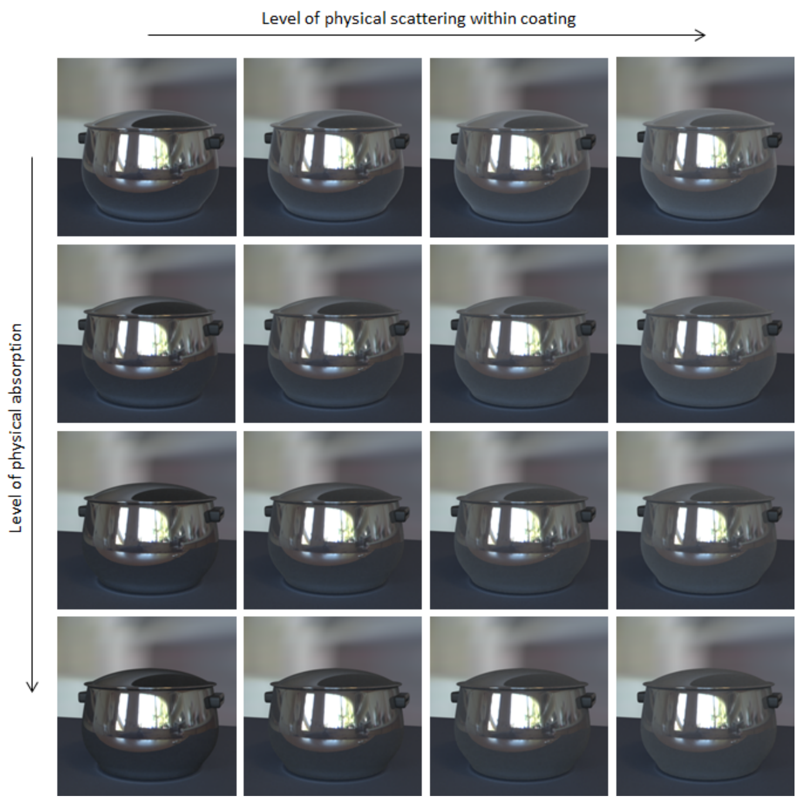

Figure 5.2: The full range of stimuli used; physical absorption varies along the y-axis, and level of physical scattering varies on the x-axis.

Research Systems). Responses were made via a standard keyboard. Observers were seated approximately 50cm away from the screen in a blacked-out cubicle, and were not required to use a chin rest, but asked to sit at a comfortable distance from the screen whilst maintaining viewing distance.

5.2.3 Statistical software

All calculations were performed using Knoblauch and Maloney's (2012) MLDS and MLCM program packages for R.

5.2.4 Observers

5 observers participated in both the MLDS experiment and the MLCM experiment for each of the three base materials (15 participants in total). All were aged 18-25 and had normal or corrected-to-normal vision.

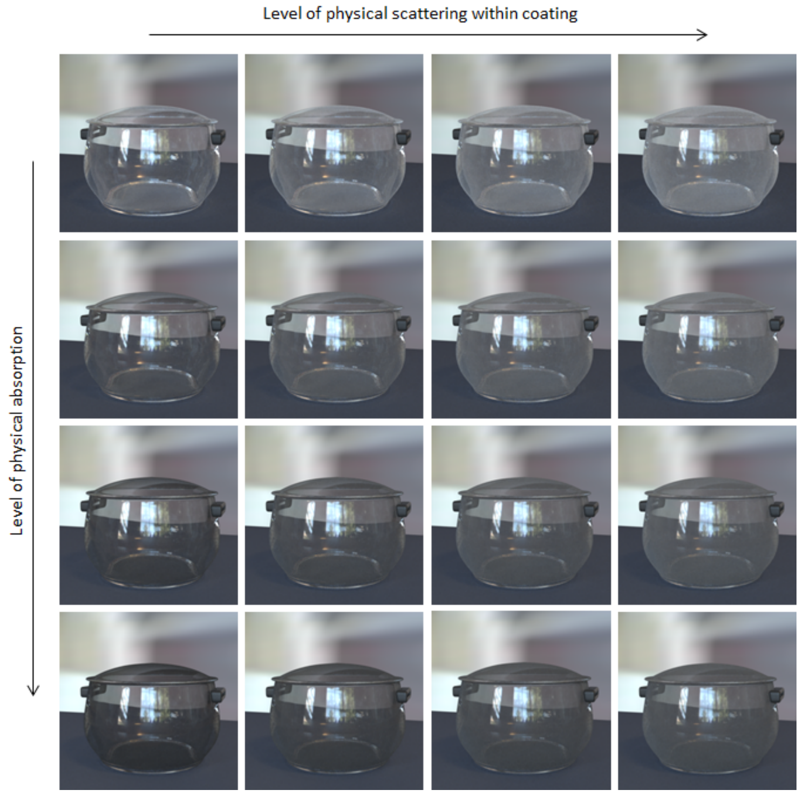

Figure 5.3: The full range of stimuli used; physical absorption varies along the y-axis, and level of physical scattering varies on the x-axis.

5.2.5 MLDS procedure

All observers completed both the absorption and scattering conditions in a single session, with a break between tasks to avoid fatigue. Completion of the two conditions by each observer was counterbalanced. The MLDS task involved presenting four images simultaneously, arranged in two pairs. Observers were asked to judge - depending on the condition - which of the two pairs had a greater difference in either level of darkness or cloudiness. The level of the parameter not being tested was kept constant at 0, to minimise potential interference. The range of values for scattering was from 0 to 36, in steps of 4 to give 10 levels, and for absorption the range was from 1 to 27 in steps of 3 to give 11 levels. For the scattering task, a total of 210 combinations was presented in a randomised order, and repeated for 3 blocks, giving 630 trials in total. For the absorption task, a total of 330 combinations were presented and repeated twice, giving a total of 660 trials. The difference in number of blocks was due to the length of time necessary to complete the task; as the full range of stimuli to be included in the MLDS testing was larger than that of the scattering task. The ordering of each pair - left to right, which had the greatest difference - was also randomised.

5.2.6 MLDS results

Results from difference scaling indicated that most observers perceived an increase in darkness or cloudiness with an increase in physical absorption and scatter, and the patterns of response were fairly consistent between observers. One observer, who completed the pyrex task, displayed seemingly random responses, suggesting that they had not understood this task, and so their data have been excluded from the study. Parameter estimates can be seen in Figures 5.4, 5.5, and 5.5. The top end of the scale for scattering in the subsequent MLCM experiment was capped at the fifth level for physical scattering, and physical absorption was capped at the seventh level. This ensured that the set of stimuli used in the MLCM tasks were approximately equally discriminable. There was a slight non-linearity at the start of the pyrex scale, so this was minimised in the scale used for MLCM by beginning at level 1 and moving in step sizes of 2.

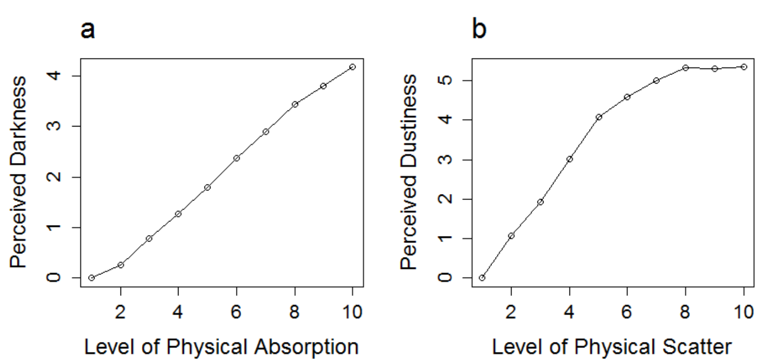

Figure 5.4: Ceramic: a) Perceived darkness (in d', on the y axis) as a function of physical absorption (x axis). b) Perceived cloudiness (in d', on the y axis) as a function of physical scatter. Both a) and b) have been normalised and averaged across all participants, and scaled to the average highest value.

5.2.7 MLCM procedure

All observers completed both the absorption and scattering conditions of the task on consecutive days to avoid fatigue. Completion of the two conditions by each observer was counterbalanced. Each condition consisted of 6 blocks of 240 trials (1440 in total), where in

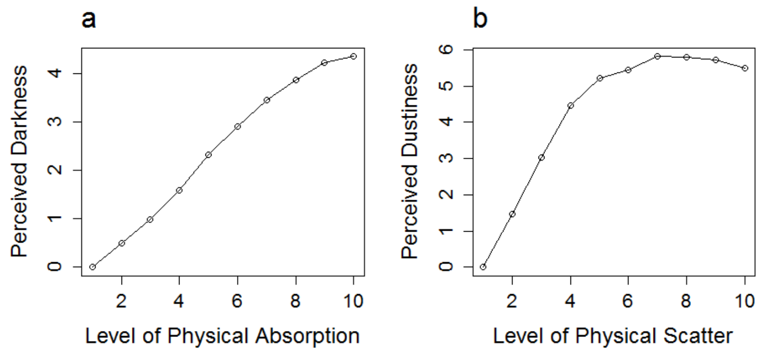

Figure 5.5: Metal: a) Perceived darkness (in d', on the y axis) as a function of physical absorption (x axis). b) Perceived cloudiness (in d', on the y axis) as a function of physical scatter. Both a) and b) have been normalised and averaged across all participants, and scaled to the average highest value.

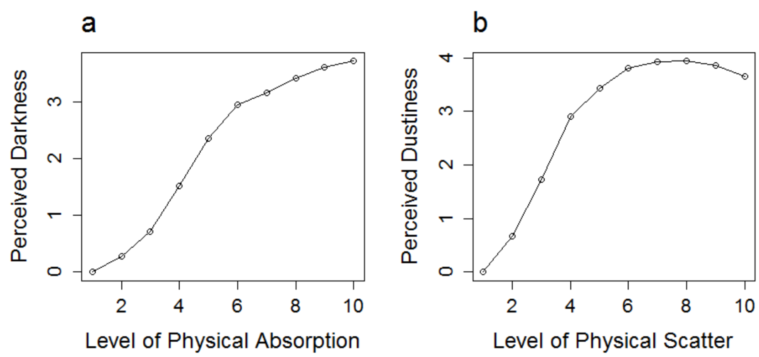

Figure 5.6: Pyrex: a) Perceived darkness (in d', on the y axis) as a function of physical absorption (x axis). b) Perceived cloudiness (in d', on the y axis) as a function of physical scatter. Both a) and b) have been normalised and averaged across all participants, and scaled to the average highest value.

an individual trial one pair of stimuli was presented on the display screen (the two images presented side-by-side). On commencing a trial, observers were presented with a black screen before a pair of stimuli was shown for a maximum of 7 seconds. This was followed by an inter-trial interval (a black screen). The observer made a judgement of whether the stimulus on the left or the right was 'darker' (higher level of absorption) or 'what you might see as dustiness or cloudiness' (higher level of scatter), regardless of the level of the other parameter, and indicated their response by pressing one of two keys. The program waited on the inter-trial interval screen for a response, and the next trial was initiated immediately upon receiving the response. Observers could make a judgement before the end of the 7 seconds given, if they did not require the full time to inspect the images. The order of trials was randomised within each block, and the ordering of each pair (left to right) was also randomised. For all materials, the range of the scattering parameter was from 4-16, in 4 levels with step size 4; the range of the absorption parameter was from 0-18 in 4 levels with step size 6.

5.2.8 MLCM results

Following the Knoblauch and Maloney method, the data were fitted to three models (saturated, additive and independent) generating three log likelihood values for each participant. A nested hypothesis test was performed on successive fits to determine the most parsimonious model for the data by calculating whether an increase in the number of parameters in the model explained a significantly higher proportion of the variance. The nested hypothesis test was an analysis of variance test on the likelihood ratios.

Ceramic

Log likelihoods and the models of best fit can be seen in Tables 5.1 and 5.2, and the parameter estimates have been plotted in units of d-prime in Figures 5.7 and 5.8. For the scatter task, three of five observers showed best fit to a saturated model, and the remaining two showed best fit to an additive model, suggesting that observers' judgements depended on a complex combination of the two variables: the estimation of the level of scattering was not independent of the level of physical absorption. Three participants showed similar patterns of response (AM, LG, and MB all perceived a decrease in cloudiness as absorption increased), however two showed different behavioural responses - CN and SK perceived an increase in cloudiness at the higher two levels of physical absorption. For the absorption task, three of five observers showed best fit for an additive model, one showed best fit to a saturated model, and one to an independent model, suggesting that they made this judgement without being as affected by the other parameter. Observers seemed better able to make judgements of darkness regardless of changes in scatter overall, and showed similar patterns of response, however physical scatter levels still affected perceived levels of darkness (but in a more constant and predictable manner).

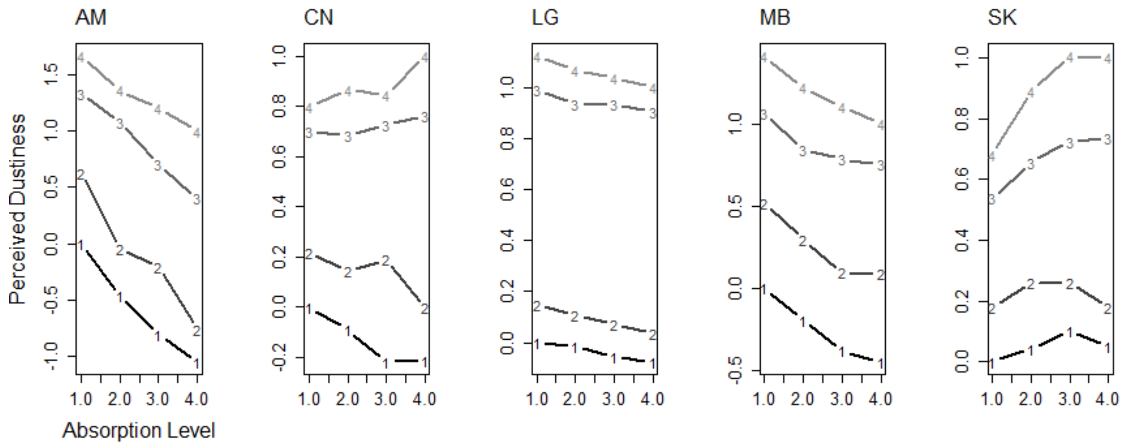

Figure 5.7: Ceramic: Perceived cloudiness for each individual (in d', on the y axis) as a function of physical absorption (x axis - numbers from 1-4 denote low to high levels of physical absorption). Numbers 1-4 within the plots denote low to high levels of physical scatter.

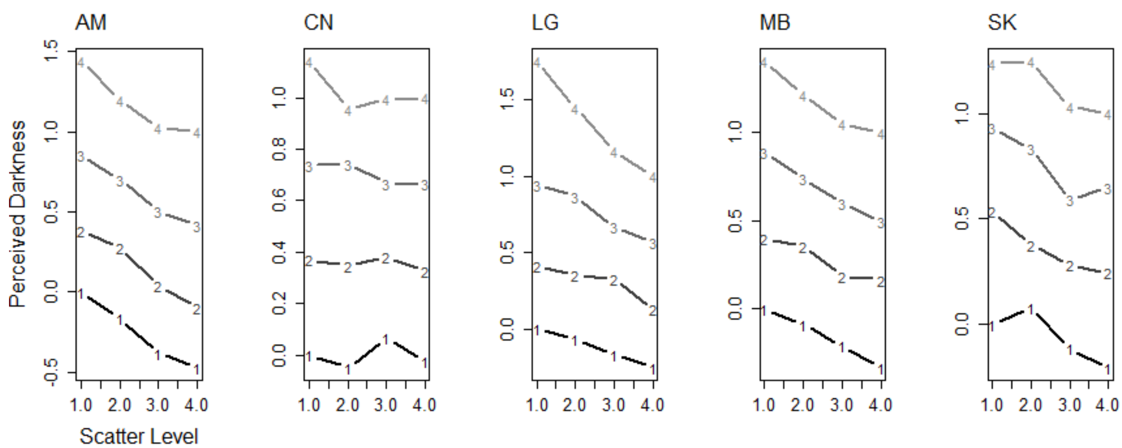

Figure 5.8: Ceramic: Perceived darkness for each individual (in d', on the y axis) as a function of physical scatter (x axis - numbers from 1-4 denote low to high levels of physical scatter). Numbers 1-4 within the plots denote low to high levels of physical absorption.

Metal

Log likelihoods and models of best fit can be seen in Tables 5.3 and 5.4, and parameter estimates have been plotted in units of d-prime in Figures 5.9 and 5.10. For the perceived cloudiness task, the model of best fit for all five observers was saturated, suggesting that their judgements of cloudiness depended on both of the variables: therefore, there was a complex interaction between the physical variables as observers did not separate the

Table 5.1: Ceramic perceived cloudiness - log likelihood values for each participant, with nested hypothesis test p-values comparing the saturated model with the additive and the additive model with the independent, and for observer CN, comparing the saturated model with the independent.

	Observer				
Model/comparative test	AM	CN	LG	MB	SK
1. Saturated model (15 parameters)	-506.972	-498.64	-312.683	-337.306	-328.97
2. Additive model (6 parameters)	-515.29	-517.802	-315.392	-340.525	353.044
i. Test: 1 vs. 2	0.055	**<0.01**	0.797	**<0.01**	**<0.01**
3. Independent model (3 parameters)	-636.007	-518.819	-358.768	-419.229	-394.402
i. Test: 2 vs. 3	**<0.01**	0.565	**<0.01**	**<0.01**	**<0.01**
ii. Test: 1 vs. 3		**<0.01**			
Model of best fit:	2	1	2	1	1

Note: numbers in bold indicate p values significant at the Bonferroni-corrected alpha level, .01

Table 5.2: Ceramic perceived darkness - log likelihood values for each participant, with nested hypothesis test p-values comparing the saturated model with the additive and the additive model with the independent.

	Observer				
Model/comparative test	AM	CN	LG	MB	SK
1. Saturated model (15 parameters)	-346.355	-383.381	-360.456	-263.290	-393.180
2. Additive model (6 parameters)	-348.286	-391.489	-373.627	-267.484	400.280
i. Test: 1 vs. 2	0.920	0.062	**<0.01**	0.495	0.115
3. Independent model (3 parameters)	-450.383	-395.705	-436.639	-354.144	-453.783
i. Test: 2 vs. 3	**<0.01**	0.038	**<0.01**	**<0.01**	**<0.01**
Model of best fit:	2	3	1	2	2

Note: numbers in bold indicate p values significant at the Bonferroni-corrected alpha level, .01

contributions of the two. There was a similar pattern of response between observers: perceived cloudiness decreased as absorption increased. However participant AP showed an increase in perceived cloudiness as absorption increased, at the higher two levels of scatter. For perceived darkness, three of five observers showed a best fit to an additive model, implying that this estimation appears to be more dependent on physical levels of absorption than on physical levels of scattering. Participant LN showed an unusual pattern of response for perceived darkness (see Figure 5.10). They perceived the high physical absorption and low scatter stimulus as very dark in comparison to all other stimuli.

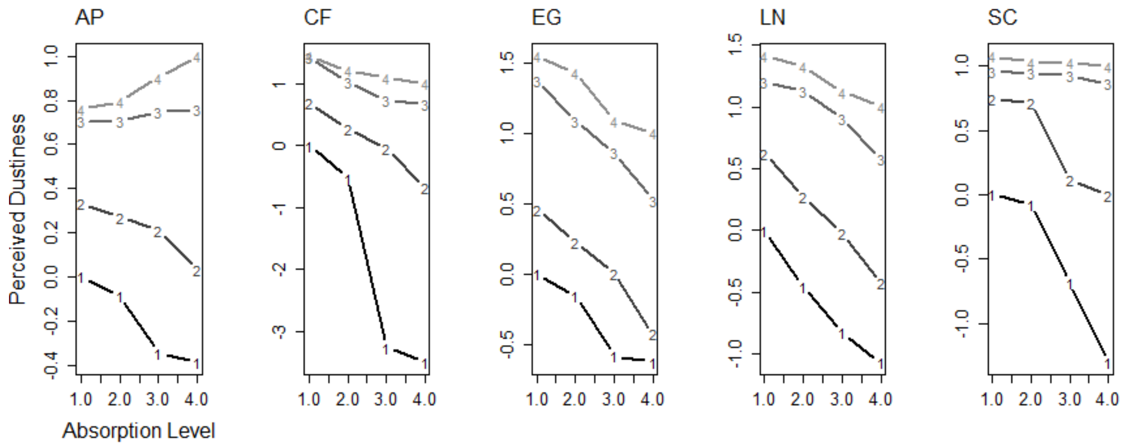

Figure 5.9: Metal: Perceived cloudiness for each individual (in d', on the y axis) as a function of physical absorption (x axis - numbers from 1-4 denote low to high levels of physical absorption). Numbers 1-4 within the plots denote low to high levels of physical scatter.

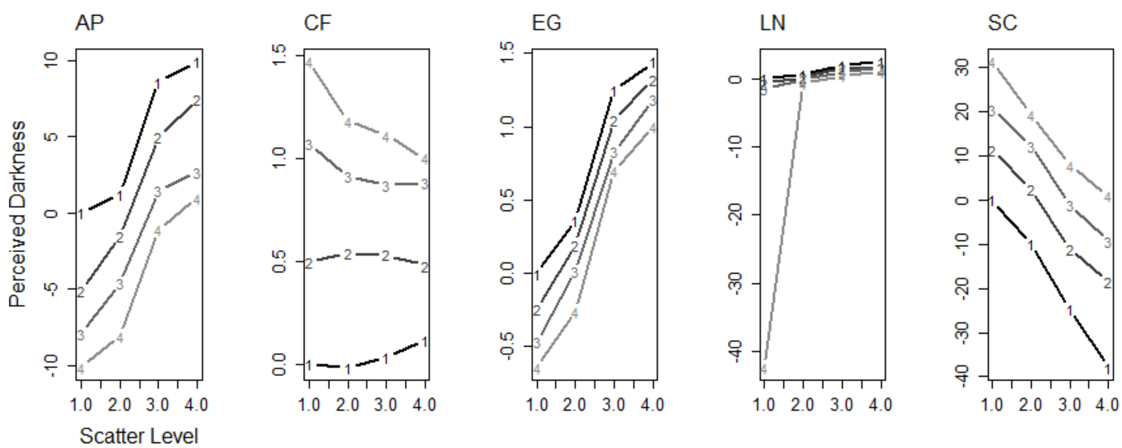

Figure 5.10: Metal: Perceived darkness for each individual (in d', on the y axis) as a function of physical scatter (x axis - numbers from 1-4 denote low to high levels of physical scatter). Numbers 1-4 within the plots denote low to high levels of physical absorption.

Pyrex

Log likelihoods and models of best fit can be seen in Tables 5.5 and 5.6, and parameter estimates plotted in units of d-prime can be seen in Figures 5.11 and 5.12. For the perceived cloudiness task, three of the four observers showed a best fit to a saturated model, and the fourth to an additive model. There are considerable differences between individuals in how the two variables contribute to their perceptions. SH perceived the stimulus with

Table 5.3: Metal perceived cloudiness - log likelihood values for each participant, with nested hypothesis test p-values comparing the saturated model with the additive and the additive model with the independent, and for observer AP, comparing the saturated model with the independent.

Model/comparative test	Observer AP	CF	EG	LN	SC
1. Saturated model (15 parameters)	-271.735	-247.389	-305.487	-216.812	-186.442
2. Additive model (6 parameters)	-358.46	-282.199	-315.232	-238.628	224.526
i. Test: 1 vs. 2	$<$**0.01**	$<$**0.01**	$<$**0.01**	$<$**0.01**	$<$**0.01**
3. Independent model (3 parameters)	-359.939	-486.259	-482.445	-433.656	-318.428
i. Test: 2 vs. 3	0.398	$<$**0.01**	$<$**0.01**	$<$**0.01**	$<$**0.01**
ii. Test: 1 vs. 3	$<$**0.01**				
Model of best fit:	1	1	1	1	1

Note: numbers in bold indicate p values significant at the Bonferroni-corrected alpha level, .005

Table 5.4: Metal perceived darkness - log likelihood values for each participant, with nested hypothesis test p-values comparing the saturated model with the additive and the additive model with the independent.

Model/comparative test	Observer AP	CF	EG	LN	SC
1. Saturated model (15 parameters)	-450.827	-506.349	-222.365	-240.187	-297.623
2. Additive model (6 parameters)	-455.453	-525.424	-226.66	-251.705	300.621
i. Test: 1 vs. 2	0.414	$<$**0.01**	0.476	$<$**0.01**	0.740
3. Independent model (3 parameters)	-850.595	-534.22	-940.476	-897.465	-754.08
i. Test: 2 vs. 3	$<$**0.01**	$<$**0.01**	$<$**0.01**	$<$**0.01**	$<$**0.01**
Model of best fit:	2	1	2	1	2

Note: numbers in bold indicate p values significant at the Bonferroni-corrected alpha level, .005

the lowest level of scatter and the highest level of absorption as considerably less cloudy than the other stimuli. Aside from that instance, SH - and LV - both broadly perceived a decrease in cloudiness as absorption increased, whereas JP and SO showed a different pattern of response for different levels of scatter. For perceived darkness, three of the four observers showed a best fit for a saturated model, and the fourth to an additive model. The two physical variables contributed in a complex way to the perceptual experience of the observers, however there was more similarity between participants in their patterns of response - perceived darkness generally decreased as physical scatter increased.

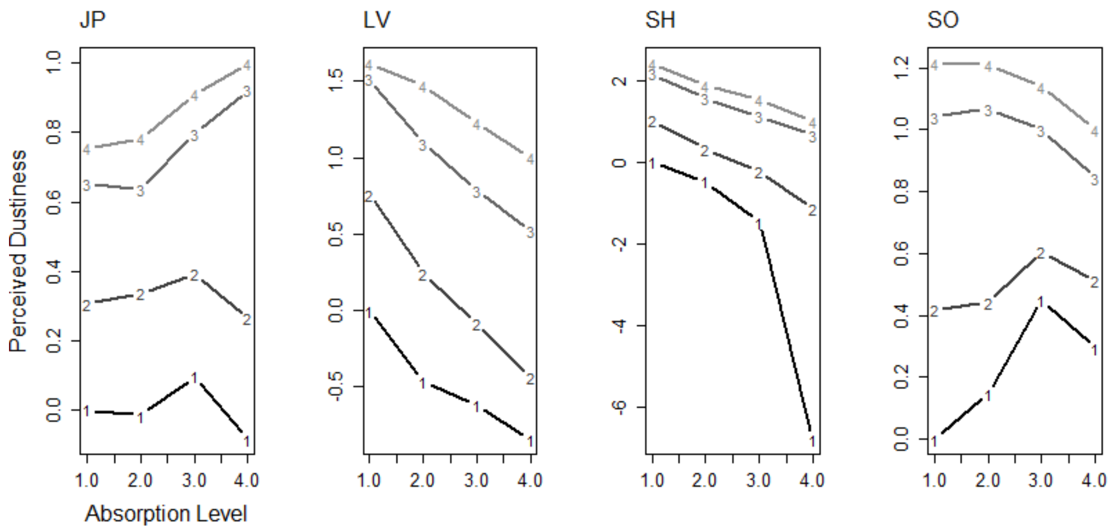

Figure 5.11: Pyrex: Perceived cloudiness for each individual (in d', on the y axis) as a function of physical absorption (x axis - numbers from 1-4 denote low to high levels of physical absorption). Numbers 1-4 within the plots denote low to high levels of physical scatter.

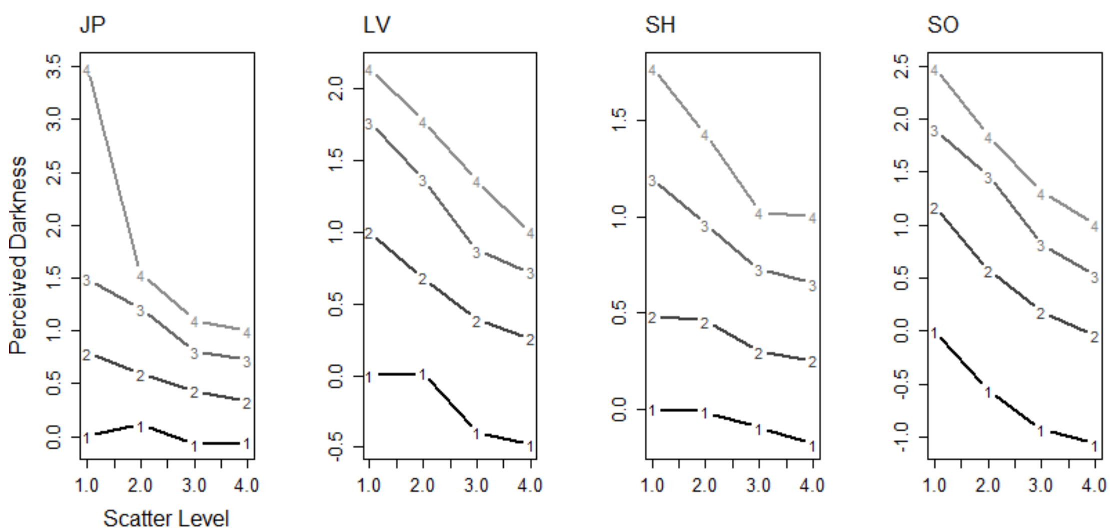

Figure 5.12: Pyrex: Perceived darkness for each individual (in d', on the y axis) as a function of physical scatter (x axis - numbers from 1-4 denote low to high levels of physical scatter). Numbers 1-4 within the plots denote low to high levels of physical absorption.

Additional analysis

In order to better compare sources of variation within and between groups, a mixed ANOVA was performed. The between subjects factor was base material, and the two

Table 5.5: Pyrex perceived cloudiness - log likelihood values for each participant, with nested hypothesis test p-values comparing the saturated model with the additive and the additive model with the independent.

	Observer			
Model/comparative test	JP	LV	SH	SO
1. Saturated model (15 parameters)	-393.281	-581.211	-315.855	-391.297
2. Additive model (6 parameters)	-425.270	-586.441	-329.434	-444.222
i. Test: 1 vs. 2	**<0.01**	0.315	**<0.01**	**<0.01**
3. Independent model (3 parameters)	-448.563	-677.533	-577.057	-460.832
i. Test: 2 vs. 3	**<0.01**	**<0.01**	**<0.01**	**<0.01**
Model of best fit:	1	2	1	1

Note: numbers in bold indicate p values significant at the Bonferroni-corrected alpha level, .01

Table 5.6: Pyrex perceived darkness - log likelihood values for each participant, with nested hypothesis test p-values comparing the saturated model with the additive and the additive model with the independent.

	Observer			
Model/comparative test	JP	LV	SH	SO
1. Saturated model (15 parameters)	-238.549	-302.070	-306.915	-276.832
2. Additive model (6 parameters)	-285.354	-317.008	-330.260	-280.990
i. Test: 1 vs. 2	**<0.01**	**<0.01**	**<0.01**	0.503
3. Independent model (3 parameters)	-433.546	-510.338	-415.210	-530.045
i. Test: 2 vs. 3	**<0.01**	**<0.01**	**<0.01**	**<0.01**
Model of best fit:	1	1	1	2

Note: numbers in bold indicate p values significant at the Bonferroni-corrected alpha level, .01

within subjects factors were physical scattering and physical absorption. The dependent variable was the parameter estimate of the observers. This was completed for each of the two tasks (perceived cloudiness and perceived darkness). For perceived cloudiness, both physical scatter and physical absorption affected perceived cloudiness ($F(3,33) = 23.708$, $p<0.001$, $F(3,33) = 9.487$, $p<0.001$ respectively), and there was an interaction between the two ($F(9,99) = 4.118$, $p<0.001$). The effect of physical scatter on perceived cloudiness did not vary with base material; however the effect of physical absorption on perceived cloudiness did differ depending on base material ($F(6,33), = 2.990$, $p = 0.019$), as did the effect of the interaction of physical scatter and absorption on perceived cloudiness ($F(18,99) = 1.999$, $p = 0.016$). Therefore, base material interacted with physical absorption, and with the combined interaction of physical scatter and absorption, to influence perceived cloudiness.

For perceived darkness, both physical scatter and physical absorption affected perceived absorption ($F(3,33) = 4.689$, $p = 0.008$, $F(3,33) = 11.244$, $p<0.001$), but without an interaction of the effect of physical scatter and absorption on perceived darkness. Perceived darkness was completely unaffected by differences in base materials.

5.3 Discussion

Overall, judgements of darkness were broadly best fit to an additive model for the metal and ceramic pots, while judgements of cloudiness were generally best fit to a saturated model. Both judgements of darkness and cloudiness for the pyrex pots were for most observers best fit to a saturated model. A best fit to a saturated model suggests that observers' perceptual estimates were dependent in a complex way on both physical variables. A best fit to an additive model suggests that observers' perceptual estimates can partially separate the influence of the two variables, but that estimates are still based on contributions of both. Results from mixed ANOVAs indicated that there were significant differences in perceived cloudiness with changes in physical absorption and in physical scatter, and that the effect of physical absorption varied different degrees depending on base material. There was also an interaction between physical scatter and absorption depending on base material. Perceived darkness changed with both physical scatter and absorption but variation in base materials did not produce any interactions.

Generally, models of interaction in perceptual judgements were more complex for judgements of cloudiness than for darkness, for ceramic and metal pots. This is consistent with previously obtained results (Chapters 2 and 3); perceptual judgements of cloudiness appear to be harder to interpret when manipulating both scatter and absorption of light, indicating a complex interacting perceptual effect. Anecdotal comments from participants also matched those from previous experiments - participants thought the darkness task harder than the cloudiness task, yet were better able to pull apart the physical dimensions when making perceptual judgements. A possible explanation in this case for perceptual judgements of cloudiness producing a more complex model might be that - as the MLDS graphs show - perceived cloudiness increased at a higher rate compared to perceived darkness. Therefore perhaps the variation in just-noticeable-differences produces more complex models of perceptual judgements for perceived cloudiness. However, these would also have an effect for the perceived absorption task as well.

For the group of participants making judgements about layers on pyrex pots, there was no difference between the levels of complexity of best-fit models for perceived darkness and cloudiness. The best fit model for perceived darkness was saturated, indicating that perceptual judgements of the layer on pyrex pots showed a diminished capacity to separate effects of absorption and scatter. We hypothesise that this might be because, although a clear material might be thought to provide the observer with additional information with which to make the perceptual decisions, the additional background information might be considered a disadvantage in these circumstances. The background against which the properties of the layer can be analysed is at a greater distance from the layer of coating than the information provided from an opaque material. Backlighting also reduces the appearance of perceptual scatter (Chapter 4, Fleming and Bülthoff 2005). Furthermore, opaque and dark materials provide a greater level of contrasting reflections, which may be particularly informative when making judgements about the properties of a surface material (Hunter, 1937). Therefore, the group of participants making judgements about darkness and cloudiness for the pyrex pots may in reality have had a smaller range of reliable pseudocues, or fewer pseudocues overall. Observers may even be better at making judgements of darkness and cloudiness with translucent volumes rather than layers, perhaps if the larger quantity of translucent material enables observers to separate information about translucence from the confounding background information (Chapter 2). With opaque base materials, the supposed 'reduction' of background information that might be employed in making perceptual judgements might instead be clarifying and enhancing existing pseudocues for the purposes of making judgements of the surface layer.

It is clear from the results we have obtained that the interaction of scattering and absorption of light in perceptual judgements is still evident. None of the results could be best fit to an independent model; reinforcing previous conclusions that perceived darkness and cloudiness cannot be predicted solely from the overall absorption or scattering of light. Furthermore, it is evident from the mixed ANOVAs that there are interactions of physical scatter and absorption with base materials in making judgements of translucence. Perceptions of layer properties therefore seem to be significantly affected by the base material to which it is applied, but only for judgements of cloudiness. This demonstrates that observers do not necessarily perceive the properties of these layers in the same way when they are coating different base materials. Responses for the pyrex group were of course very different from the ceramic and metal groups, but the results from these latter two also produced different results (sufficiently different between groups, rather than within) to suggest that the layers are still not perceived in the same way. This might be attributed to the mirror-effect of the metal pans perhaps providing the information expected to be advantageous with the pyrex pots, but in a more useable or informative way - whereas the ceramic pots reflect only one colour with reduced information about reflections in the scene, which limits the cues available on the surface of the object. However, the results suggest that best-fit models for the metal pots were closer to saturated interactions than the ceramic - so perhaps the consistency in base material beneath the layer of coating aided perceptual judgements; as potentially conflicting cues may have been minimised.

Preface to Chapter 6

The first four experimental chapters reported have primarily focused on the ways in which translucent materials are perceived. In the real world the surfaces of objects are rarely spatially uniform. Natural surface layers may be inherently non-homogenous, may become damaged or may be the result of random accumulation of dust and dirt. In the next chapter, we begin to address how non-homogeneous translucent layers might influence perception of shine. In addition, we investigate how behavioural decisions can be influenced by abstract concepts such as cleanliness (which in some contexts might be equated with shine).

Chapter 6

Criteria for judging 'shine' or 'cleanliness' with varying subsurface light scatter

Gloss and translucence are important perceptual properties of many materials. The physical determinants of translucence and gloss are similar, as both depend on the scattering of light, either from a surface or in propagation of light through a material. Previous studies have demonstrated that layers of translucent material that vary in subsurface scattering, enveloping base materials, can be used as a manipulation of gloss. However when these layers of scattering material are found in day to day life, they are rarely completely uniform. This experiment set out to investigate how sensitive human observers are to changes in variation of scattering across a surface, and how such changes might influence perceived gloss. As changes in variation of a scattering layer also alter the thickness of the layer at any given point on the surface, we varied layer thickness as well as the coefficient of variation of scattering of a layer in a maximum likelihood conjoint measurement task to assess whether any effect of variability was independent of changes in the overall thickness of a layer. Observers were asked to make a two-alternative forced-choice judge-

ment of shine. In addition, glossy objects in many situations might also be described by observers as clean. We ran the same task with a second group of observers who were asked to make judgements of cleanliness, to see if the perceptual estimates might vary differently when the same given physical variables were manipulated. We found that both increases of variation in scattering and layer thickness significantly affected perceived shine; increases in variation contributed to judgements of decreasing shine and cleanliness to a lesser degree than layer thickness. Variation of scatter was discovered to contribute significantly differently for judgements of shine compared to cleanliness. The implications of these results for the weighting of cues in internal perceptual models are discussed, as well as the implications regarding perceived shine.

6.1 Introduction

Translucence is an important property of many materials. Changes in translucence greatly affect the perceived realism of natural materials, and discrepancies can result in the 'uncanny valley' effect. Human waxworks may appear to be very realistic at first, but as the wax does not have the same translucence properties as real skin it does not quite look like real human skin, and observers feel a sense of unease (Mori & Kageki, 2012). Translucence is a result of volume transport of light: that is, light travelling beneath the surface and being scattered from within a material. Scattering of light within volumes of material produces perceptual translucence, and also when there is scattering of light in smaller layers just beneath the surface of an object. Many natural materials, including human skin, have such layers of translucent materials. However, changes in scattering in layers of a material do not just alter perceived translucence, but can also change perceived shine or roughness (Chapter 4.5). The physical basis of roughness and translucence are almost identical. Physical shine or roughness is essentially manipulated by the scattering of light from the surface of a material. The more light is scattered, the rougher a material appears. The only difference is that translucence is perceived when scattering can be seen within the volume. Human vision is clearly sensitive to variation in layers of scattering medium. For example, without a layer of scattering medium to simulate fine hairs on rendered skin, the resulting image looks significantly less realistic (Koenderink & Pont, 2003). In previous chapters, experiments have investigated how our perceptions of these kinds of translucent layers are influenced by contextual factors, but how sensitive are we to variation within these layers themselves?

Another kind of layer with this characteristic is that of dust on the surface of objects. In the real world, such layers do not simply vary in thickness or composition, but in uniformity as well - e.g. dirt or dust on crockery, or general wear and tear eroding a coating on an object (Beekman, den Harder, Viergever, and van Rijk 1997; Jacquemoud and Ustin 2008 and Donner, Weyrich, d'Eon, Ramamoorthi, and Rusinkiewicz 2008 also illustrate the non-uniformities of light scattering in layers of natural materials, namely plant matter and human skin). In this experiment we set out to investigate the importance of irregularities in layers of scattering materials when making judgements of perceived shine, and as variation in uniformity also changes the minimum and maximum thickness of a layer, whether the effects of change in uniformity were independent of the effects of variation in thickness.

As already outlined, changes in translucence in surface layers can simulate changes in shine (Chapter 4.5). Rather than a rough surface scattering light and preventing specular reflection, scattering of light beneath the surface of a material diffuses the appearance of highlights, contrast gloss, and distinctness of reflected images, all of which influence perceived gloss. In everyday life, we might also commonly judge very shiny objects to be clean, as changes in surface texture - such as imperfections and blemishes - reduce perceived gloss (Chadwick & Kentridge, 2015). Do we use the same cues for judging cleanliness as for gloss and shine, when the perceptual cues varied are identical? No previous studies have attempted to answer this question. Furthermore, comparisons of these distinct judgements would allow us to determine whether perception of a material property goes beyond simple estimation of physical properties and can be influenced by related - but more abstract - knowledge about the concept of cleanliness. It might be conjectured that when making judgements of cleanliness we equate this concept with our perception of shine for objects with a smooth surface, but this is not necessarily the case. Therefore, we wanted to investigate two additional questions in this experiment: whether an increase in perceived glossiness equates to perceived cleanliness, given the manipulated variables, and whether changes in irregularity and maximum thickness (as a control for irregularity changing maximum thickness) contribute in the same way to judgements of gloss and cleanliness.

The second hypothesis is essentially asking whether perceptual judgements or decisions might be altered or weighted differently according to the question asked. It begins to question where in the chain of visual processing the perceptual properties might interact when making decisions. Properties can interact at the physical level, such that the interactions are tangled within the visual cues available to the visual system, and they may also interact in the perceptual model of the observer - either at the detection, or visual processing stage, or at the cognitive and conceptual stage of making decisions. This question therefore mirrors a well-established finding in colour vision, where very different colour matching results may be obtained by asking whether observers think the hues match, or whether two patches have been cut from the same piece of paper (Arend, Reeves, Schrillo, & Goldstein, 1991).

To investigate these questions, stimuli that varied in both average thickness and the coefficient of variation of scatter in the layer of translucent coating were generated. A preliminary maximum likelihood difference scaling (MLDS; Knoblauch and Maloney 2012) experiment was used to produce a set of stimuli with approximately equal perceptual differences in each of the two variables. A maximum likelihood conjoint measurement (MLCM; Knoblauch and Maloney 2012) task was used, where groups of observers were asked to make decisions about shine or cleanliness. The data were analysed with the MLCM method, by fitting three alternative models, each varying in the number of estimated parameters (Knoblauch and Maloney 2012, and Maloney and Yang 2003). We did not expect that either physical parameter would map precisely onto either perceptual quality tested. Instead, we aimed to investigate what the relative contribution of the two variables might be towards perceptual judgements of glossiness or cleanliness.

6.2 Method and Results

6.2.1 Stimuli

A metal pot was simulated using Blender and Luxrender: software which models objects and produces images through raytracing in a physically realistic way, respectively. A layer of coating material which scatters light uniformly was applied to the pot. The overall average thickness of the layer was manipulated, as well as the coefficient of the variation of the thickness of the scattering layer (the units of these variables are arbitrary). The original stimulus set comprised 720 renders images, including 12 levels of average thickness (ranging from 0.00005 to 0.0006 in steps of 0.000005) and 12 levels of the coefficient of variation (ranging from 0.05 to 0.6, in steps of 0.005). At lower levels of layer thickness, the pans appear only mildly dusty, whereas at high levels they appear coated in a thick film of dust. As the coefficient of variation of scattering increases, the spatial arrangement of the pattern of scattering stays the same but the levels of scatter at each point reach greater extremes around the average thickness of the scattering layer. Therefore, at higher values of coefficient of variation, there is a greater contrast between patches of lower and higher scattering.

Maximum likelihood difference scaling (MLDS; Knoblauch and Maloney 2012) was performed with the stimuli. MLDS aims to produce a perceptually uniform scale of stimuli. Observers were presented with a two alternative forced choice task, asking them to compare two pairs of stimuli which varied in only one physical dimension (either the degree of uniformity, or layer thickness). As two pairs of stimuli are presented at once, five different versions of the pattern in variation within the scattering layer were produced for each combination of the two variables, so that no two simultaneously presented images had the same spatial pattern of non-uniformities. Two separate MLDS pilots were carried out, one for each physical dimension. This established a range within the set of stimuli where each step made through the set was approximately equally discriminable, to produce an approximately linear perceptual scale. This range was 0.15-0.3 for variation, and 0.00005-0.0002 for thickness (see Figure 6.1).

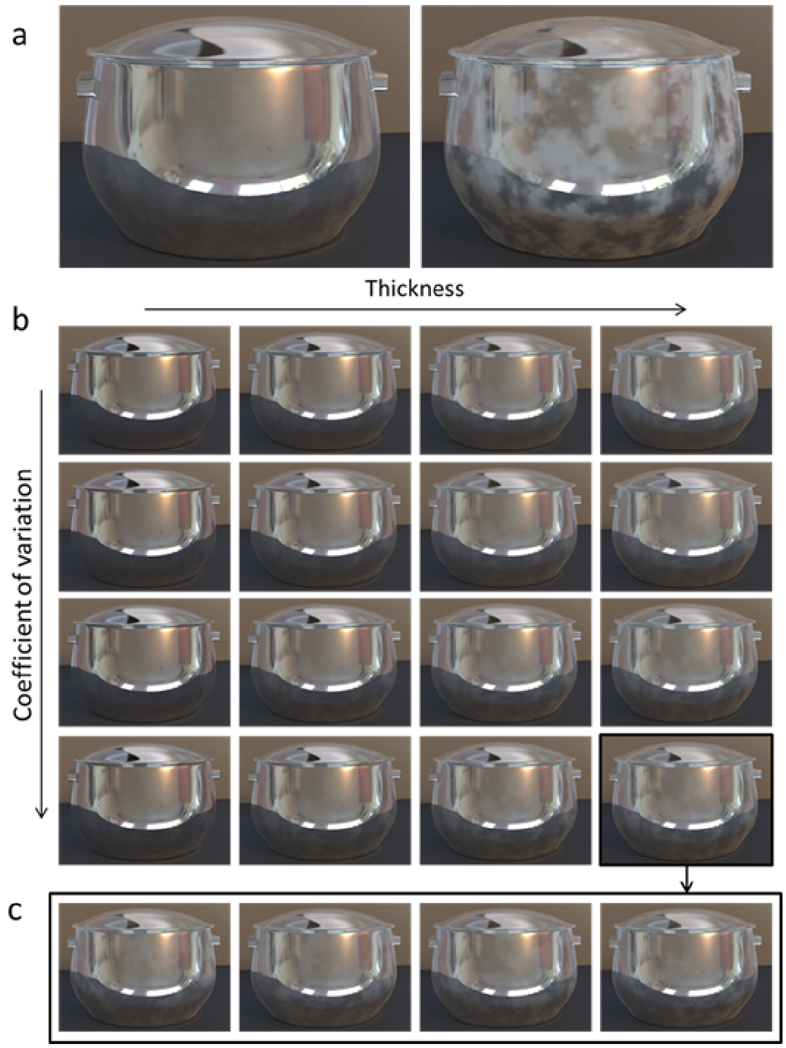

Figure 6.1: a) illustrates the difference between a stimulus with low variation in the distribution of scatter and with a thinner layer of scattering, and a stimulus with high variation in the distribution of scatter and a thick layer of scattering. b) illustrates the full range of stimuli used in the experiment, in one of the five patterns of noise. c) demonstrates the four alternative noise patterns used.

6.2.2 Apparatus

Experimental software was written in MatLab. Stimuli were presented on a gamma-corrected ViewSonic 17 display monitor (1064x768 pixels, with refresh rate of 100Hz), controlled by means of a CRS Visage (Cambridge Research Systems). Responses were made via a standard keyboard. Observers were seated approximately 40cm away from the screen in a blacked-out cubicle, and were not required to use a chin rest, but asked to sit at a comfortable distance from the screen whilst maintaining viewing distance.

6.2.3 Statistical software

All calculations were performed using Knoblauch and Maloney's MLDS and MLCM packages for R (2012).

6.2.4 Observers

Two observers completed the MLDS task, and two groups of five observers took part in the MLCM tasks.

6.2.5 MLDS

In order to obtain approximately linear spaces for the two physical variables, two participants completed the MLDS tasks. One physical parameter is held constant, while the other is varied; two pairs of stimuli are presented in a quadrant on the screen. Observers are then asked to decide which of the two pairs shows the greatest perceptual difference. Each participant was therefore asked to judge how thick a layer of 'dust' was on pots varying only in the thickness of the layer, and on a different task asked to judge the difference in 'blotchiness' of the layer of 'dust' on the pots. The order of completion of the two tasks was counterbalanced. For each task, a total of 300 trials were presented - with 6 steps through the full stimulus set made (with a step size of 2 levels of the original scale), giving 20 blocks of 15 combinations for each task. A relatively small number of combinations of stimuli were produced, given the number of steps made through the stimulus set, so the number of blocks was increased to compensate for this to obtain enough trials for use with MLDS analysis.

6.2.6 MLDS results

Figure 6.2a shows the parameter estimates for perceived blotchiness. The linear section of the plot between levels 1 and 2 of physical variation in the distribution of scatter demonstrate that observers could not distinguish the stimuli at those levels. Therefore, for the MLCM task, the base level of the coefficient of variation was set to the third level used in MLDS. Figure 6.2 shows the parameter estimates for perceived thickness of the layer. As the level of physical layer thickness increases beyond the fourth level, the perceived difference between levels decreases, showing that observers were less able to distinguish stimuli at these levels. The base level of layer thickness was therefore set to the first level used in MLDS. Four steps were made through the stimuli of size 1 (0.000005, and 0.005, for thickness and coefficient of variation respectively).

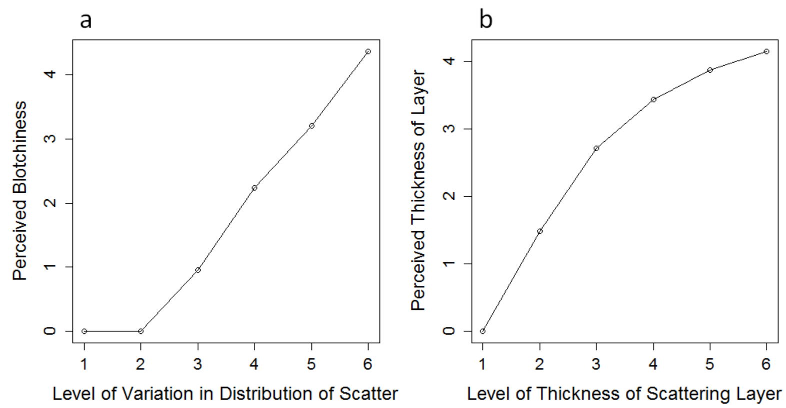

Figure 6.2: a) Perceived 'blotchiness' (in d', on the y axis) as a function of the level of variation in the distribution of physical scatter (x axis). b) Perceived thickness of layer (in d', on the y axis) as a function of the level of thickness of the scattering layer. Both a) and b) have been normalised and averaged across all participants, and scaled to the average highest value.

6.2.7 MLCM

Maximum likelihood conjoint measurement sets out to determine the contribution of two separate variables towards a single perceptual decision. Observers are presented with pairs of stimuli, while two different variables are manipulated simultaneously. Observers are asked to make a single judgement about the pairs - for instance, which of the two looks shinier. The patterns of response obtained for each observer can then be used to calculate how the observer's responses vary as a function of the two different variables. MLCM analysis then compares these results with a number of different models, which vary in the number of hypothetical parameters, and finds the model which best fits the data. In MLCM, three different models are calculated: saturated, additive, and independent. The saturated model (which allows for fifteen hypothetical parameters) assumes that the values of both variables are needed in order to predict an observer's response, and therefore by allowing for complex interactions makes no assumption of linearity. The additive model (which allows for six parameters) assumes that there is some degree of linearity in the interaction, and therefore no complex interaction between the two variables in the perceptual estimates. An independent model (which allows for three parameters) assumes that there is no interaction of the two variables in the perceptual estimates.

In this experiment, two separate groups of five observers were given an MLCM task, each group completing only one of the two tasks. One group was asked to make perceptual judgements of shine, and the other was asked to make judgements of cleanliness. Each task comprised 1200 trials, with 5 blocks of 240 combinations.

6.2.8 MLCM results

Data were fitted to three models using the MLCM package - saturated, additive, and independent - producing three log likelihood values for each participant. A nested hypothesis test was performed on successive fits to determine the most parsimonious model for the data by calculating whether an increase in the number of parameters in the model explained a significantly higher proportion of the variance.

Results from one participant in the shine group were excluded, as their responses indicated that they were not answering the question or had not understood the task.

Log likelihoods and nested hypothesis test values can be seen in Tables 6.1 and 6.2. It is evident that for both tasks, the two physical parameters contributed in a generally additive manner towards perceptions of shine or cleanliness. For the cleanliness task, four of the five observers' results were best fit by an additive model, and a saturated model provided the best fit for the remaining observer. For the shine task, all four observers were best fit by an additive model. As layer thickness increases, perception of shine and cleanliness decreases, and as variation in 'blotchiness' increases, perception of shine and cleanliness again decreases.

While individual differences might seem to indicate some inconsistency between participants, the graphs of the patterns of responses for each individual show very minor differences in judgements - the overall trend appears to be fairly consistent (see Figures 6.3 and 6.4). The points which appear to shift the model of best fit from additive to saturated may be reflecting individual differences with particular combinations of physical parameter levels. The graph for participant AM in Figure 6.4 shows a broadly similar pattern of responses to the other observers, with perceived cleanliness generally decreasing as blotchiness increases - with the exception of the combinations of the third and fourth levels of blotchiness and the first level of layer thickness. At the third level of blotchiness, AM appears to give the response that this stimulus looks cleaner than the stimulus at the second level; however at level four, perceived cleanliness decreases again (but not far below the rating of the second level). Therefore, we conclude that while the model of best fit for this participant was saturated rather than additive, these differences do not necessarily show broader inconsistencies between individuals.

A preliminary analysis of slopes using a t-test was performed for each participant on the additive models, and increases in layer thickness (M = 1.011, SD = 0.306) were found to contribute more to judgements of shine and cleanliness than increases in coefficient of variation (M = 0.215, SD = 0.196; $t(8) = 6.540$, $p<0.001$). However, the coefficient of variation seemed to have a slightly higher contribution to overall decisions of perceived cleanliness than for perceived shine. That is, for equal increases in the coefficient of variation of the scattering layer, perceived cleanliness decreased more than perceived shine. Therefore a mixed ANOVA was performed, with two within subjects factors (type of physical variation (TV) - 2 levels, thickness or coefficient of variation - and strength (ST), with 4 levels), and one between subject factor of task question (QU, 2 levels - shine or cleanliness). The dependent variable was the parameter estimate. Effects of thickness and variability on parameter estimates were different (TV: $F(1,7) = 3.399$, $p<0.01$), and effects of different levels of thickness and variability on parameter estimates were significantly different, regardless of the between-subjects task (ST: $F(3,21) = 76.200$, $p<0.001$). Furthermore, different levels of thickness and of variability affected parameter estimates differently (TV by ST: $F(3,21) = 66.687$, $p<0.001$). Crucially, this also differs significantly between the two task groups (shine and cleanliness). That is, differential effects of thickness and variability also differ between the shine and cleanliness tasks (QU by ST by TV: $F(3,21) = 4.038$, $p<0.021$). This can be seen in Figure 6.5, which illustrates the contributions of layer thickness and coefficient of variation towards judgements of shine (Figure 6.5a) and cleanliness (Figure 6.5b). The slope of the coefficient of variation is steeper for perceived cleanliness than for perceived shine, demonstrating that as the coefficient of variation increased, perceived cleanliness decreased more than perceived shine.

Table 6.1: Perceived shine - log likelihood values for each participant, with nested hypothesis test p-values comparing the saturated model with the additive and the additive model with the independent.

Model/comparative test	Observer EG	HR	SH	YB
1. Saturated model (15 parameters)	-338.533	-478.646	-302.550	-414.060
2. Additive model (6 parameters)	-345.255	-485.982	-309.466	-420.392
i. Test: 1 vs. 2	0.144	0.100	0.128	0.178
3. Independent model (3 parameters)	-362.521	-492.350	-340.011	-434.328
i. Test: 2 vs. 3	**<0.01**	**<0.01**	**<0.01**	**<0.01**
Model of best fit:	2	2	2	2

Note: numbers in bold indicate p values significant at the Bonferroni-corrected alpha level, .0125

Table 6.2: Perceived cleanliness - log likelihood values for each participant, with nested hypothesis test p-values comparing the saturated model with the additive and the additive model with the independent.

Model/comparative test	Observer AM	BAR	EL	LS	SW
1. Saturated model (15 parameters)	-495.959	-578.061	-360.245	-427.605	-298.081
2. Additive model (6 parameters)	-507.619	-583.473	-369.073	-431.289	302.092
i. Test: 1 vs. 2	**<0.01**	0.288	0.039	0.599	0.532
3. Independent model (3 parameters)	-573.371	-618.084	-398.293	-488.339	-413.762
i. Test: 2 vs. 3	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**
Model of best fit:	1	2	2	2	2

Note: numbers in bold indicate p values significant at the Bonferroni-corrected alpha level, .01

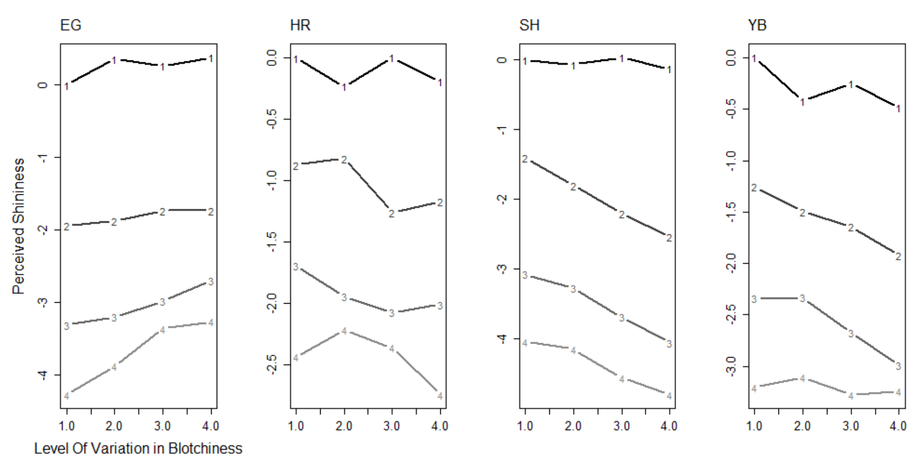

Figure 6.3: Perceived shininess (on the y axis) as a function of the level of variation in blotchiness (x axis - numbers 1-4 indicate low to high levels of variation). Numbers 1-4 within the plots denote low to high levels of thickness of the scattering layer.

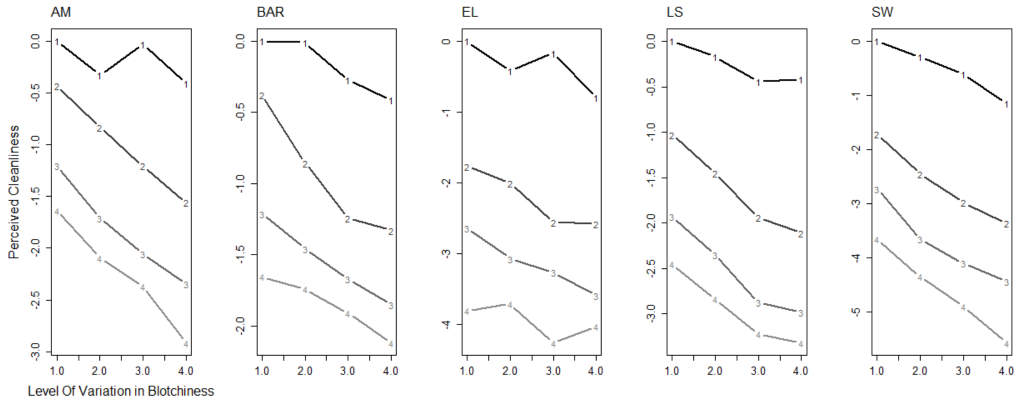

Figure 6.4: Perceived cleanliness (y axis) as a function of the level of variation in scattering (x axis - numbers 1 to 4 indicate low to high levels of variation in scattering). Numbers 1-4 within the plots denote low to high levels of thickness of the scattering layer.

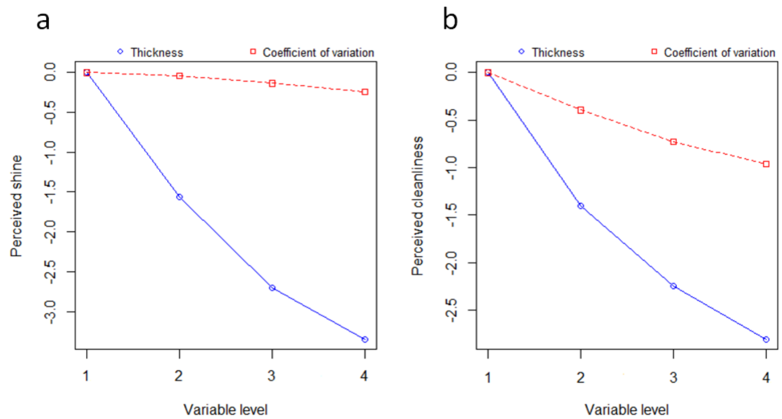

Figure 6.5: Additive models illustrating the contributions of layer thickness and coefficient of variation towards a) perceived shine and b) perceived cleanliness, each averaged across observers.

6.3 Discussion

In this experiment we investigated explicitly how variations in a layer of scattering coating ('blotchiness' and thickness) might contribute to judgements of overall gloss. We did not assume that either of the two physical dimensions manipulated could be mapped directly onto perceived shininess or perceived cleanliness.

With an increase in either of the two parameters alone perceived shininess decreased, and when both parameters were manipulated an additive effect was produced. Overall perceived shine decreased more with higher levels of both parameters. There was a slightly higher contribution towards perceived shine from the overall average thickness of the layer of scattering for all participants.

We were also interested in relating perceived shine to the concept of cleanliness. The results were very similar for perceived cleanliness as for perceived shine. Increasing each parameter individually resulted in a decrease in perceived cleanliness, and there was generally an additive effect of manipulating both parameters together. Interestingly, there was a higher contribution of the coefficient of variation of scatter ('blotchiness') towards judgements of perceived cleanliness compared to judgements of perceived shine, and different levels of thickness and variability affected parameter estimates differently. This demonstrates that the criteria for perceptual judgements were weighted differently, depending on the decision to be made. This does not necessarily mean that the stimuli themselves were perceived differently, as appropriate ranges were chosen within MLDS testing, and it is unlikely that the two groups perceived the stimuli in a significantly different way as each participant was randomly assigned.

This potentially provides evidence for a model of perception which allows differential weightings of cues when making different (but similar) perceptual judgements. For perceived shine, participants weight various cues differently and make judgements based on a number of cues or pseudocues. Each individual might interpret these cues slightly differently, or weight or prioritise them in a different way, as there are between-observer differences as well - this finding is well established (Chadwick & Kentridge, 2015). For perceived cleanliness, the model of perceived shine might well feed into or supply the means for the decision, or the same pseudocues may be employed, but we seem to apply a secondary level of judgement. This could be at a cognitive level since judgements of cleanliness are not wholly based on perception, but require knowledge of what it means to be clean. This could redistribute the weightings for the pseudocues to reassess the perceptual input. Differences between the weightings of perceptual cues in this experiment would also imply that we cannot equate shininess and cleanliness in all situations. And of course, in other situations it may not be the two physical parameters manipulated here that are the primary source of information to the observer.

Preface to Chapter 7

The previous experiment demonstrated how, by manipulating translucence of layers of coating to vary the physical gloss of a surface, perceived shine was influenced by both the average level of physical scattering and also the variation in scatter across the surface.

In the studies so far conducted, several different words have been used to describe observers' perceptual decisions - including 'darkness', 'dustiness', 'cloudiness', 'shine', and 'cleanliness'. It is known that the question asked of observers can alter the responses obtained, just as with the difference between perceived shine and cleanliness as in the previous chapter. It is worth noting that the labels chosen in the experimental tasks may have affected the patterns of data obtained. Asking a slightly different question would affect the responses obtained for those experiments, but terms were specifically chosen to be the relevant 'opposite' term for e.g. perceived shine for the object or scenario in that experiment. We have shown that responses do change with different questions like 'shine' and 'clean', but these differences are also subtle, so while a particular term might be debated for its relevance the use of a different term would not yield startlingly different findings. As is now clear, the labels used to describe the perceptual judgements of the observers are therefore not the same as the manipulated variables. These manipulated variables could be considered in the context of a generative model: the physical variables are the generative variables, which the observers do not explicitly consider. If we frame it thus, the data obtained in the previous experimental chapters could also be interpreted as a measure of the extent to which both generative variables contribute to an observer's internal model of the meaning of the labels.

In the next study, we take a multi-sensory approach to the perception of gloss. When making real world judgements of objects and materials, we do not only have access to visual information but also tactile information. Here an experiment is reported investigating how perceived gloss is influenced by haptic feedback - namely, variation in fine scale texture.

Chapter 7

Visual judgements of gloss are influenced by tactile roughness

Human observers are adept at visually perceiving changes in many surface properties including colour, shape and texture. The study of gloss perception is a rapidly expanding avenue of research as an offshoot of the study of perceived texture, and we now have a better idea of factors that are involved in perceiving gloss. However, in everyday life, we do not merely interact with objects and materials by looking at them, but also by touching them. Tactile feedback from a surface potentially provides a great deal of information about the properties of a material, and several studies have begun investigating the integration of tactile feedback with visual experience. A recent study discovered that varying haptic friction significantly influenced perceived gloss: both produced main effects on perceived gloss when varied with visual gloss, and friction also interacted with visual gloss (an increase in friction paired with an increase in visual gloss countered one another in perceptual judgements of gloss - Adams et al. 2016). However, physical roughness is one of the main tactile properties observers use to categorise materials when making tactile judgements of objects. We believe that the simulation of haptic friction with a Phantom robot guiding the observer's fingers could not adequately simulate

this haptic physical property. In this study, we set out to explore whether variation in surface roughness influenced perceived gloss when presented alongside visual representations varying in shine. Observers took part in a simple rating task, where on a given trial a single real glass with a particular surface roughness would be presented behind a mirror for observers to touch, while a rendered visual image of the glass (of a particular level of shine) was also presented so that the location of the real glass and virtual image coincided. Observers rated the gloss of the glass as an object (as opposed to separately rating visual or tactile gloss). Both tactile and visual gloss were varied from trial to trial. Perceived gloss was significantly influenced by both visual and tactile gloss, but for the majority of observers there was no interaction in the contribution of the two variables. This suggests that while tactile surface roughness does influence perceived gloss, it does not interact with visual gloss when observers are making judgements of perceived gloss, as was found with haptic friction. The implications of these results are discussed in relation to the findings of Adams et al. and to potential physiological differences in processing of tactile sensations.

7.1 Introduction

How do people integrate the 'feel' of an object and the way it looks in making judgements about a surface? Attempts have been made to determine how visual and haptic senses compare in visual material perception tasks. Baumgartner, Wiebel, and Gegenfurtner (2013) conducted an extensive exploratory investigation, identifying how well participants could categorise a wide range of materials based on either visual or tactile cues. Categorisation performance was less consistent with haptic than with visual information. However ratings were highly correlated between the two, and material samples were similarly organised within the perceptual space for both senses. The principal components identified for distinguishing materials were hardness and roughness. Furthermore, in a later study (Baumgartner, Wiebel, & Gegenfurtner, 2015) it was shown that such haptic material representations could emerge independently from visual experience; vision did not assist haptic learning, and even the lack of vision did not boost haptic characterisation. In recent years, studies have shown that human observers combine tactile and visual cues when making decisions about size and slant (Ernst & Banks, 2002; Ernst, Banks, & Bülthoff, 2000). Some evidence from fMRI studies seems to identify neural integration of visual and haptic information in relation to texture - significant haptic and texture selective responses in areas near those recruited in visual texture discrimination, whereas those same areas were not recruited for haptic shape processing (Podrebarac, Goodale, & Snow, 2014). In addition, looking at textures seems to evoke activity in neural areas associated with tactile stimulation, potentially showing an expectation of contact with surface for the purpose of planning hand grip and placement (Sun, Welchman, Chang, & Di Luca, 2016). A number of papers have explored the relationship between tactile and visual input in relation to perceived surface gloss. Kerrigan, Adams, and Graf (2010) showed that haptic cues affected perceived gloss by manipulating perceived compliance and highlight displacement. When hard and smooth, objects were thought of as shiny for larger highlight displacements than when soft and rubbery, where objects continued to appear matte with smaller offsets between highlight and diffuse shading. More recently, Adams et al. (2016) further demonstrated that touch modulated perceived gloss, by using a Phantom - a haptic device which simulates sensations by applying force feedback to a user's hand - to simulate different compliances (soft and rubbery, to hard) and frictions (from 'slippy', or low friction, to high friction). Compliance did not influence perceived gloss, while friction interacted with perceived gloss. Observers could easily detect increases in gloss when paired with low friction, but high gloss with high friction produced a small perceptual change. They concluded that the visual system treats visual gloss and haptic friction as correlated cues to surface material properties.

While these findings represent a significant breakthrough in the understanding of multisensory judgements of material properties, the kinds of tactile information manipulated do not encompass all the variation potentially encountered in surfaces varying between matte and glossy. Surface roughness (on a fine-scale level of texture - where the surface geometry is defined at a lower resolution than microscale, as with coarse sandpaper) was the second of the two most important components identified by Baumgartner et al. (2013)

in distinguishing materials on the basis of haptic information. The measures used in the study by Adams et al. (2016) appear to comprise mainly compliance (the hardness of a surface - the other important component identified by Baumgartner et al. 2013) and friction, in terms of 'slipperiness' compared to high friction. As simulated by the Phantom, we think that while haptic friction is undoubtedly important, it does not necessarily include roughness information. There are several different types of mechanoreceptor, each specialised to receive different kinds of tactile information (Johansson & Flanagan, 2009; Purves et al., 2008). For example, the bulbous corpuscles detect tension deeper within the layers of skin, such as higher levels of pressure, whereas Meissner corpuscles are sensitive to changes in fine-scale texture and light touch. It is possible that these two cues are interpreted by separate physiological mechanisms, and processed differently. We therefore investigated whether this kind of tactile information has a similar level of influence over perceived gloss.

Observers were presented with images of glasses ranging from matte to glossy in a mirror-screen apparatus, in which a mirror reflected the stimuli images from a high resolution screen to the observer so that the stimuli appeared to be in front of the observer. The mirror obscured the view of the real tactile stimuli, which varied in fine-scale texture from very smooth to very rough, that were placed in front of observers, and so observers were told that they would be touching the object they saw in the mirror. They were asked to rate the glossiness of the glass on a scale from 1-10, on the basis of how they looked and felt.

7.2 Materials and Methods

7.2.1 Visual stimuli

Images of glossy glasses were produced using Blender, an open source 3D computer graphics program which can simulate and model 3D scenes and objects. Images were then rendered using LuxRender, a ray-tracing renderer. LuxRender simulates physical properties of materials, including their light-absorbing, -transmitting and -scattering properties. It is based on PBRT (Physically Based Ray Tracing, Pharr and Humphreys 2004) and simulates the propagation of light through the scene in a physically realistic way.

The original stimulus set comprised 8 rendered images. The glossiness of the glass was manipulated by altering the UV roughness component for the glossy surface material in the model, which alters the sharpness of reflections. A real-world lighting probe employing natural spectral light distributions illuminated the scene. A structured box shape, simulated in a matte wooden material with a cut-out, was included in the background to provide cues to the angle of the presentation of the glass. A stand held the glass at a 45 degree angle leaning away from the perspective of the observer, so that the sides of the glass (and highlights/surface roughness) would be more visible to the observers. This stand was invisible to the observer from the perspective of the camera, but a real physical version was used to hold the real glasses in place.

Maximum likelihood difference scaling (MLDS) was performed with the initial set of stimuli, with one participant. This technique (Knoblauch & Maloney, 2012) measures how perceived material properties vary as a function of a physical scale, to estimate how changes in a single physical dimension translate to perceived difference estimates. Two pairs of stimuli differing in one dimension are presented simultaneously (where for the four dimension values of the stimuli a, b, c and d, $a<b<c<d$), and the observer is asked to decide which of the two pairs shows a greater perceptual difference. Multiple comparisons are made with all possible combinations of the stimulus set. These responses are then used to calculate a maximum likelihood perceptual scale, which relates the perceptual parameter estimates to the physical scale. These calculations were performed with the MLDS package for R. The resulting function relating physical scale to parameter estimates was then used to calculate estimated values of the physical variable which would produce a perceptually constant scale for perceived gloss when manipulating roughness on a simulated glossy material. A new set of 10 stimuli was then created using this estimated perceptually constant scale. This was done with the intention of removing larger non-linearities in the perceptual scale, so that any genuine interactions between different sources of information are not caused by non-linearities in the stimulus space. MLDS was performed for a second time on the second set of stimuli. Figure 7.1a shows the perceptual scaling with the first set of stimuli, and Figure 7.1b shows the perceptual scaling for the second set of stimuli, illustrating how the perceptual differences between levels are more consistent in the new set. The values of roughness used for the images post-MLDS were 0.8, 0.53755, 0.44815,

0.38147, 0.32586, 0.27716, 0.23328, 0.19303, 0.15564, and 0.12057 (which produce images of glasses ranging from matte, through low- to high-gloss - see Figure 7.2). After initial piloting of the experimental task, it was discovered that combining the extreme levels of the visual and tactile stimuli (e.g. the roughest tactile glass, and the shiniest visual stimulus) was experienced as so unrealistic that observers did not know how to make a judgement. Therefore, the two extremes of the visual stimuli - the most matte, and most glossy - were removed from the experimental task, and used as ‘anchors’ for the observers’ scale of response.

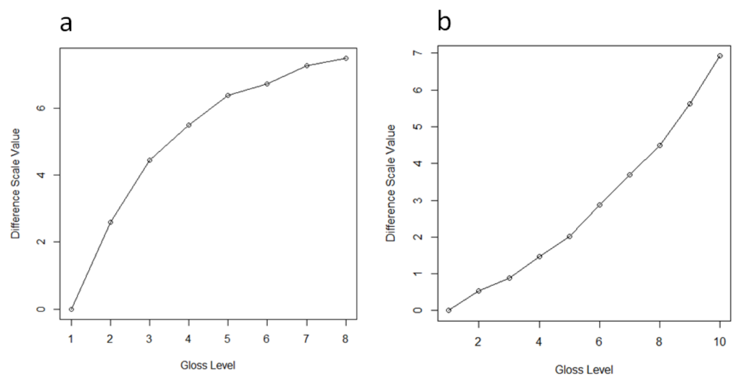

Figure 7.1: a) shows perceived shine, in difference scale value on the y axis, as a function of the gloss level of the stimuli (arbitrary unit) before performing difference scaling. b) shows perceived shine (y axis) as a function of gloss level (arbitrary units), with the second set of stimuli created following maximum likelihood difference scaling.

7.2.2 Tactile stimuli

Four tactile stimuli were produced, ranging from rough and matte to very smooth. Glass tumblers were coated with either matte paint, a 2/3 matte and 1/3 silk mixture of paint, a 1/3 matte and 2/3 silk mixture of paint, or silk paint, using an air spray. Once dry and cleaned from residual powder, a small number of naïve observers were asked to compare them and order them in terms of roughness/smoothness. The majority were able to do so satisfactorily with some minor confusion between the middle two glasses, however smaller differences at either ends of the scale were not detectable. The glasses were measured with a Diavite DH-7, a high precision surface roughness meter which calculates *Ra* (in micrometres, μ), the average of the absolute values in a roughness profile, the most com-

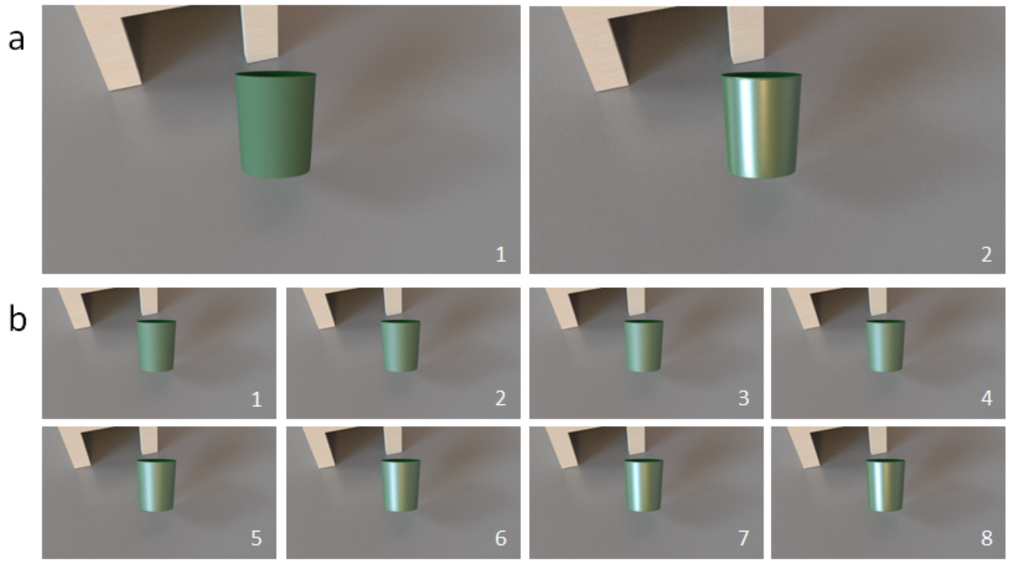

Figure 7.2: a) The anchor stimuli shown prior to the experiment - the stimuli at the two extremes of the scale of glossy glasses produced. b) The eight stimuli used in the experiment, from low to high gloss.

monly used one dimensional roughness parameter which gives a good general description of height variations in a surface. Three measurements were taken for each level of tactile roughness. The average Ra for the matte surface was 7.03μ, 2.71μ for ⅔ matte ⅓ silk, 1.97μ for ⅓ matte ⅔ silk, and 1.11μ for silk. None of the individual measurements taken for each glass overlapped with the range of measurements of the other glasses.

7.2.3 Observers

One participant completed the MLDS task. Eight observers took part in the experimental task. All were aged 18-30 and had normal or corrected to normal vision.

7.2.4 Apparatus

Experimental software was written in Matlab. Observers were seated in front of the mirror-screen apparatus, where visual stimuli were presented on a Samsung D550 full HD plasma screen (51", a full resolution of 1920x1080 pixels, and a refresh rate of 60Hz), and reflected to the observer in a mirror such that the glasses in the reflection appeared to be on the table in front of the observers. Tactile stimuli were placed in front of observers, underneath the mirror, on a small stand which held the glasses at a 45-degree angle leaning away from the observers. The visual and tactile stimuli were positioned such that when touching the tactile stimuli, it appeared to observers that they were touching the same glass that appeared on the mirror (see Figure 7.3 for an illustration). Observers were asked to use a chin rest and an eyepatch covering one eye, as the images were not presented in stereo to observers, and therefore gave only monocular cues for depth.

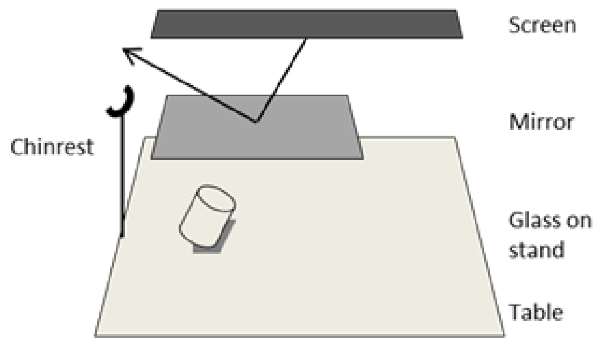

Figure 7.3: The mirror-screen apparatus used. Observers, when seated and using the chin rest, could only see images reflected in the mirror, and not the real glasses beneath the mirror.

7.2.5 Procedure

In a single trial, a single visual stimulus was presented on the screen for 3 seconds, followed by a black screen. Just before each trial, the experimenter placed a tactile stimulus on the stand beneath the mirror. Participants were asked to look at the visual stimulus and touch the tactile stimulus simultaneously, and to stop touching the tactile stimulus as soon as the visual stimulus disappeared from the screen. All participants were asked to touch the stimulus lightly, so as to ascertain the fine-scale texture, rather than rub the glass with unnecessary pressure. Participants were then asked to give a rating of perceived gloss of the glass they were touching and looking at on a scale from 1-10, where 1 indicated completely matte and 10 indicated high gloss. The experimenter recorded the rating, and the next trial began. All combinations of the 4 tactile and 8 visual stimuli were used, giving 32 trials presented randomly in each block, and repeated for 10 blocks in a session. Participants completed two sessions, on different days to avoid fatigue, giving a total of 640 trials completed.

7.3 Results

An analysis of variance was performed, with two within subjects factors of visual gloss and tactile gloss, and estimated gloss as the dependent variable. Mauchly's test indicated that the assumption of sphericity had been violated, and so degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity. Main effects of both visual gloss and tactile gloss were found ($F(1.103, 7.719)=34.306$, $p<0.01$, $F(1.272, 8.901)=31.519$, $p<0.01$). No interaction was found between visual and tactile gloss ($F(3.319, 23.235)=2.104$, $p = 0.122$). Both visual and tactile gloss had a significant effect on perceived gloss, but there was no interaction in their contribution. Figure 7.4 shows the results of each individual separately - there are clear individual differences in patterns of response, demonstrating that decision-making criteria varied considerably between observers, although generally both tactile and visual gloss increased perceived gloss ratings. Some observers prioritised visual gloss over tactile feedback, and for others tactile gloss had a larger effect on perceived gloss ratings. When tactile stimuli were compared directly observers could differentiate between them, however when presented in the task alongside the visual stimuli there was some uncertainty over the level of tactile gloss as the lines representing different levels of physical roughness in Figure 7.4 are not always cleanly separated. Individual analyses of variance showed that three of the eight observers demonstrated an interaction between visual and tactile gloss (EL: $F(21,639) = 9.35$, $p<0.01$, PM: $F(21, 639) = 1.59$, $p = 0.046$, YB: $F(21,639) = 1.77$, $p = 0.019$) in addition to the main effects found with the rest of the observers.

7.4 Discussion

This study evaluates the way in which visual and tactile cues to gloss are combined when people have access to both in making a decision about the glossiness of an object. Glossiness is treated as a property of the object rather than a specifically visual or tactile quality. All observers showed statistically significant main effects of both tactile and visual stimulus variables. This is consistent with cue integration occurring. The results, however, show considerable individual differences in the extent to which these stimulus properties influence observers' judgements. Three observers were found to have a statistically signif-

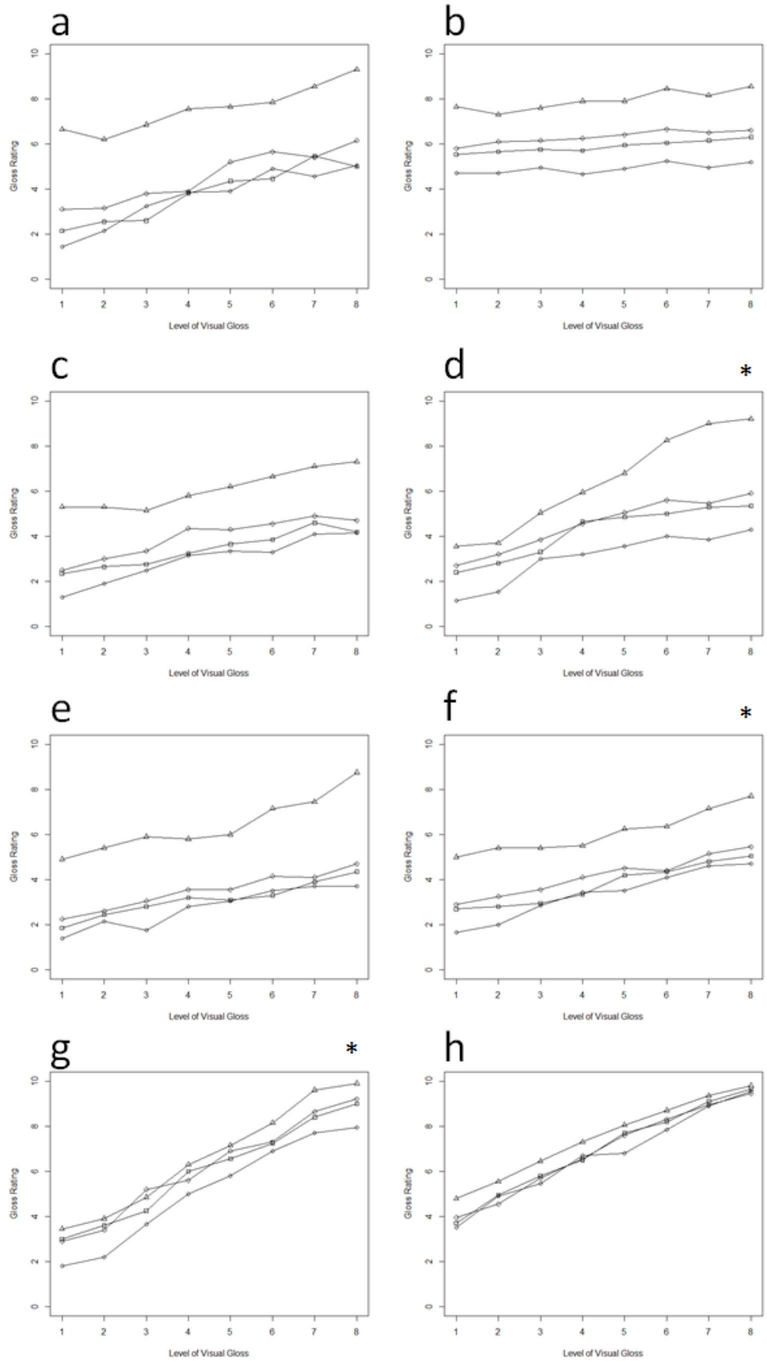

Figure 7.4: Perceived gloss (rating from 1-10) as a function of levels of visual gloss (x-axis) and level of tactile roughness/smoothness (multiple lines - where 1 = matte, 2 = 66% matte, 3 = 66% gloss, 4 = gloss), for each individual. Individual graphs marked with '*' showed a significant interaction.

icant interaction between visual and tactile gloss as well as main effects, whereas all other observers showed only main effects of the two variables. It seems that individual observers were prioritising different kinds of information as patterns of response were not all the same, even within the group that showed an interaction and within the group that did not. Of those observers whose results showed an interaction, one observer (Figure 7.4d) mainly showed a higher rate of increase in rated gloss for the highest combined levels of both visual and tactile gloss. The ratings of another observer (Figure 7.4g) at the two lowest and two highest levels of visual gloss were very similar, with smaller differences in ratings between levels of tactile gloss. In the group of observers whose results did not demonstrate an interaction, some were influenced primarily by visual gloss (Figures 7.4a, e, h) with little separation of the first three tactile levels, but with a much larger increase in rated gloss for the highest level of tactile gloss. A third observer's decisions appear to be influenced more by tactile rather than visual gloss (Figure 7.4b), as there was good separation of the four lines but relatively little change with increases in visual gloss along the x axis. This suggests that observers may be combining sensory information in different ways, either at the level of sensory integration or at the level of cognitive decision-making, depending on how important, or perhaps, how congruent they judge the information to be. Several observers' decisions were dominated by one or other of the two variables - perhaps the apparent conflict between the two kinds of information in the task meant that they ultimately relied on only one of the two to make their decisions.

It has been shown in studies such as those by Ernst et al. (2000) and Ernst and Banks (2002) that the reliability of cues has a significant effect on cue integration. In the case of haptic and visual cue integration, Ernst and Banks concluded that the aim of integration was to minimise variance in the final estimate of the object property (be it shape or size), and so visual cues are prioritised when there is more variation with estimations of tactile properties. When both cues provide fairly reliable information and there is an overlap between the probability distributions of the two estimates they are integrated and weighted according to their variance, but when the distributions do not overlap then the cue with the least variance is prioritised - so whether there is an interaction might relate to how good observers are at processing the individual cues. Changes in cue integration are a probable explanation of what happened during pilot tests with the lowest and highest levels of visual gloss (that were subsequently used as the visual anchors). Here the tactile and visual information may have been so incongruent in the observers' perceptual experience that they were unable to combine them. Even for those observers where an interaction between the two variables was found, this was not consistent between participants and there was no clear overall pattern.

Another possible reason for variation between participants might be that they were not all equally capable of making judgements of fine scale texture, even if they could discriminate between them when presented simultaneously. This would reflect the findings of Baumgartner et al., who found that tactile feedback generally correlated with judgements of visual gloss, but that discrimination alone was not as reliable when only using tactile information compared to using just visual information. This would explain why many observers seemed to prioritise visual over tactile information, if they were assuming that the tactile information they were receiving was less reliable than their visual experience.

The overall results obtained here are generally consistent with those of Adams et al.. In their study, it was found that visual gloss cues and haptic friction cues were integrated. This is consistent with the additive effects found in the results presented here. In the rating experiment conducted by Adams et al., there was also a significant interaction term, which we did not find as consistently in our results. One might have expected to find a strong and consistent interaction of tactile roughness with visual gloss, as with haptic friction. This may well be due to statistical power differences, however there is also a possibility that perhaps the kinds of tactile stimuli used are of fundamentally different types, in that haptic friction cannot be equated to surface roughness. One concern in interpreting the results presented above might be that the levels of roughness in the tactile stimuli were insufficiently discriminable. It is unlikely that this had any significant effect however, as participants were able to order the tactile stimuli reliably. Although in principle a help, difference scaling of the tactile stimuli would be almost impossible to perform, as not only would the task take too long for a participant to complete, but it would be impossible to make stimuli across a sufficiently wide range of roughnesses using our current methods. Furthermore, if some of the stimuli were made rougher than those in the current set they would feel more incongruous and unconvincing when presented with shiny visual stimuli, and so judgements would become more difficult to make and less useful in testing hypotheses regarding cue interactions and integration. In piloting, observers reported being very aware that the difference between the shiniest visual stimulus and the roughest tactile stimulus was too conflicting, and that they did not consider a judgement would be realistic (they felt they were overthinking the decision based on the two kinds of information). It may therefore be the case that observers are not as good at detecting small differences in roughness and that when uncertainties arise they generally resort to using primarily visual information.

It is also possible that the particular method used here limited the level of cue integration found. Observers reached out to touch the glass beneath the mirror while viewing a rendered image, and therefore did not see their own hand touching the glass. Models of causal inference in multi-sensory perception (e.g. (Körding, 2007)) would suggest that the lack of a visible hand during the trial might decrease observers' ability to infer that the two cues are combined. In initial pilot runs of our experiment, we attempted to address this potential issue by asking participants to observe the rendered image and then touch the real glass beneath the mirror when the rendered image presented had been changed to show a box completely covering the rendered glass. Observers reported that this felt more incongruent than simply touching the glass while viewing the rendered image, perhaps as they could still not see their arm reaching inside the box, so this step was removed. In future experiments, alternative methods could be explored, such as increasing the translucence of the reflective mirror, or attaching lights to the observer's hand so that it is visible through the mirror when they reach for the glass.

These results have a number of other implications. If roughness as a 'type' of tactile information is being interpreted in different ways, or combined with visual information in a different way, this suggests that the information is being processed separately. In the experiment by Adams et al., there was a main effect of friction, and an interaction effect of friction with visual gloss on perceived gloss. Perhaps these different kinds of tactile information combine with visual information differently; again, this may be determined by the reliability of judgements for each type of sensory information. We might be more certain that we are not being fooled by our senses when we feel high or low friction, compared to changes in roughness or compliance (which had no main effect on perceived gloss). It has been shown that perceived 'squishiness' (compliance) can vary greatly depending on whether observers are looking at the object that they are touching (Di Luca, 2016), suggesting that *perceived* reliability of perceived compliance is higher when visual feedback is available. It might be conjectured that information from different kinds of mechanoreceptor are processed separately, or combined differently with other sensory modalities such as vision. To explore and confirm this hypothesis, it would be interesting to manipulate all these different cues in a single experiment to investigate how they are combined.

While the results reported here are consistent with an interpretation of a level of successful cue integration, the results are also consistent with interpretations in terms of cue switching ((Adams, 2016)) or response bias. A cross-modal adaptation experiment would enable us to explore the effects found in this experiment further, investigate which of the possible interpretations such as cue integration or cue switching are more plausible, and confirm whether the hypothesis regarding varying degrees of combination of different tactile information is tenable. Participants would be asked to repeatedly touch materials with a given level of roughness in an aim to adapt them to that degree of tactile roughness, and would then make visual judgements of surfaces varying in roughness, experienced in conjunction with a constant tactile stimulus. If there were adaptation effects on judgement of visual roughness that depended on the roughness of the tactile adaptor, this would confirm that perceptual processing of tactile roughness influenced visual judgements of the surface. This method could then be repeated with friction and compliance as in Adams et al., to see if there was a difference in the way the kinds of tactile information combine with visual information.

Summary

In this experiment, we found that varying both visual gloss by means of physically accurate rendered object images, and tactile gloss by means of glasses varying in surface roughness, significantly influenced ratings of perceived gloss. As visual gloss increased, perceived gloss increased, and as tactile roughness increased, perceived gloss decreased. We also found that for the majority of observers tested there was no significant interaction effect of tactile and visual gloss on perceived gloss. This is interesting because previous findings from Adams et al. showed - similarly to our results - a main effectw of haptic friction on perceived gloss, but also found an interaction between haptic friction and visual gloss. The lack of interaction of surface roughness (for the majority of observers) with visual gloss in our findings suggests that the tactile properties of haptic friction and roughness of materials are either detected by different physiological mechanisms (such as the bulbous corpuscles, or Meissner corpuscles) or that the two kinds of sensory cues are processed separately. Previous findings from Baumgartner et al. (2013) demonstrated that observers are less reliable at discriminating materials by haptic roughness alone, suggesting that - if less capable of discriminating materials - observers may place less importance on the information they perceive in this modality. Cues to perceived gloss from different sources seem to be integrated in different ways, depending on perceived congruity or reliability of the observers' ability to perceive a cue accurately.

Preface to Chapter 8

The experimental chapters have so far explored the ways in which translucent materials are perceived, and how translucence may relate to gloss. Gloss and translucence are visual properties of surfaces. Are the perceptions of gloss and translucence processed by the same pathways which process other visual surface properties, like texture and colour? A previous study with a neuropsychological patient demonstrated that perceived gloss was not reliant on areas responsible for colour and texture. This final experimental chapter reports a study of a neuropsychological patient which set out to test whether the cortical processing of translucence might be reliant on areas found to be responsible for texture or colour.

Chapter 8

Translucence perception is not dependent on cortical areas critical for processing colour or texture

Translucence is an important property of natural materials, and human observers are adept at perceiving changes in translucence. The perception of different material properties do not arise from the same cortical regions, and it is therefore plausible that the perception of translucence is dependent on specialised regions, separate from those important for colour and texture processing. We tested MS, a cortically colour blind observer, who performs at chance on tasks of colour and texture discrimination. In addition to reassessing his performance on the Farnsworth-Munsell 100 Hue test, we tested MS with two translucence ranking tasks. In the first task, stimuli were images of glasses of tea varying in tea strength, in the second stimuli were glasses of tea varying only in milkiness. MS was able to systematically rank both strength and milkiness, although less consistently than controls. An additional group of controls tested with greyscale versions of the images still succeeded at the tasks, although they performed less consistently on the milkiness task, demonstrating that some cues to translucence perception do not rely on colour information.

The systematic performance of MS suggests that some aspects of translucence perception do not depend on regions critical for colour and texture processing.

8.1 Introduction

Visual identification of materials requires the ability to visually discriminate a range of properties such as colour and texture. Many natural materials not only reflect light from their surfaces but are also translucent. Translucency is an optical characteristic which is caused by light scattering below the surface of an object, that is, it is scattered within the material. The two most important physical parameters that determine how light is transported within a material are the scattering and absorption coefficients, which essentially capture the rate at which light spreads and is attenuated as it travels through the material. Making judgements of the purity and concentration of mixtures of translucent materials (e.g. how strong tea is and how milky) is not simply a matter of identifying a material as being translucent, but also of estimating its scattering and absorption parameters. Human observers are capable of identifying materials within fractions of a second (Sharan et al., 2009), and can readily discriminate visually between very similar translucent materials such as skin or fruit and their waxwork replicas, despite the subtlety of such differences. Many of these fine differences are thought to be due to the differing properties of translucent layers at the surface of materials. As we appear to be so proficient at detecting differences in translucence (Jensen et al. 2001, Vasseleu, in Cubitt, Palmer, and Tkacz 2015, p.163-178, and Murakoshi, Masuda, Utsumi, Tsubota, and Wada 2013), and translucence seems to be associated with important properties of natural materials such as the health of skin or ripeness of fruit (Fleming & Bülthoff, 2005; Fleming, Jensen, & Bülthoff, 2004; Hetherington, Martin, MacDougall, Langley, & Bratchell, 1990), it is plausible to ask whether there is a region of cerebral cortex specialised for processing translucence.

The properties that contribute to material identification are not all processed at the same neural locus. Cant and Goodale (2007) first demonstrated that attending to material properties or shapes of objects activates distinct regions of cortex. Cavina-Pratesi et al. (2010a; 2010b) went on to show that texture and colour discrimination rely on distinct areas of cerebral cortex - posterior collateral sulcus vs. anterior collateral sulcus and lingual gyrus - and in between the two there is a region activated by both colour and texture. Kentridge, Thomson and Heywood (2012) tested neuropsychological patient MS, who lacked the areas implicated in processing texture and colour and found that he could, nevertheless, discriminate glossiness independently of variation in lightness or texture. Perceived glossiness must therefore be processed by an area of cortex distinct from those areas. As we have noted, these properties - colour, texture, and gloss - are not all we need to discriminate materials: many objects are either translucent or have a translucent layer at the surface, which we can clearly distinguish. For instance, human skin has a very distinctive translucency, as do fruit flesh, marble, plants, plastics, minerals, and foods such as meat, cheese, and liquids, e.g milk. It is the particular way that the volume transport of light through the substance interacts with the material, reflected back to the observer, which produces this translucency. Translucence is also conceptually distinct from material properties such as colour, gloss and texture. Physical translucence depends on the degree to which light is scattered and absorbed within a material. Perceived translucence appears to depend on pseudocues (cues which the visual system extracts from the scene, which act as heuristics rather than deterministic cues to the physical properties being perceived - Chadwick and Kentridge 2015) which are indirectly driven by these physical changes, and that can also encapsulate complex interactions of light transport within the material (Chapter 2). Factors affecting perceived translucence include the amount of light absorption (darkness), direction of illumination, the shape and size of the object, and colour (Fleming, 2014; Fleming & Bülthoff, 2005; Fleming, Jensen, & Bülthoff, 2004).

In order to investigate whether translucence of materials might be processed in a region distinct from colour and texture, we tested neuropsychological patient MS along with a small group of control participants (one age-matched and two non-matched). Following brain damage as a result of suspected viral encephalitis in 1970, MS developed dense achromatopsia (colour blindness of cortical origin) accompanied by a left hemianopia (with macular sparing), prosopagnosia and visual object agnosia (Heywood and Kentridge 2003; Kentridge, Heywood, & Cowey, 2004); however, he has intact Snellen acuity (Mollon, Newcombe, Polden, & Ratcliff, 1980). He performs at chance on tasks of colour and texture discrimination, but above chance on tasks of glossiness discrimination (Cavina-

Pratesi et al. 2010a, Kentridge et al. 2012). Neuroimaging indicated that MS lacks the regions normally activated by texture and colour in healthy observers (Cavina-Pratesi et al., 2010a, 2010b). Structural MRI (Heywood, Cowey, & Newcombe, 1991) revealed extensive bilateral lesions to ventromedial occipito-temporal cortex, damage to the 2nd, 3rd, 4th, and 5th temporal gyri in the right hemisphere, and damage to the right temporal pole. Right striate cortex was completely destroyed, which accounted for the left field hemianopia. In the left hemisphere, there was damage to the left temporal lobe, confined to the left temporal pole, 4th temporal gyrus, and hippocampal gyrus. There was also bilateral ventral-occipital damage to the lingual and fusiform gyri.

To test MS's ability to perceive translucence, we presented him with a number of tasks, using translucent stimuli which varied in their absorption and scattering of light - photographs of glasses of milky tea. We asked MS to rank order stimuli on the basis of strength of tea (primarily dependent on light absorption), and milkiness (primarily dependent on light scattering). If MS could discriminate such properties it would imply that elements of perceived translucence are processed in a region of cortex distinct from those processing colour and texture which he lacks. All stimuli used were images of real tea (creation of which is detailed in the Supplementary Materials). If MS failed at this task, it would suggest that he was unable to distinguish volumes on the basis of changes in translucence or absorption.

Initially, we determined whether MS understood what it meant for a liquid to look more or less absorbing or cloudy. He described the liquids as either "looking more like beer than water" (indicating a liquid with a stronger concentration of tea), and agreed that some appeared more like milk compared with water (indicating a liquid of high milk concentration).

8.2 MS's colour vision

We established that MS's condition remained unchanged by asking him to complete the Farnsworth-Munsell 100 Hue Test, a task in which he was asked to arrange a number of equiluminant chips in chromatic order. In an early test, MS had scored 1245 (Mollon et al., 1980) - Figure 8.1a shows the hue chips as ordered by MS, and Figure 8.1b illustrates the test scoring. His total error score was 1268 and confirms that MS has not recovered any colour vision in over three decades. A score would be expected to be in the range of 170-195 for a normal age-matched observer (Kinnear & Sahraie, 2002) - the worst 5% of performances in the normal population score between 80 and 195, depending on age. Previous performances of achromatopsic observers have resulted in a mean score of 582 (Bouvier & Engel, 2006). However, MS's score reflects random responding, which would correspond to a score of approximately 1200 (Victor, 1988).

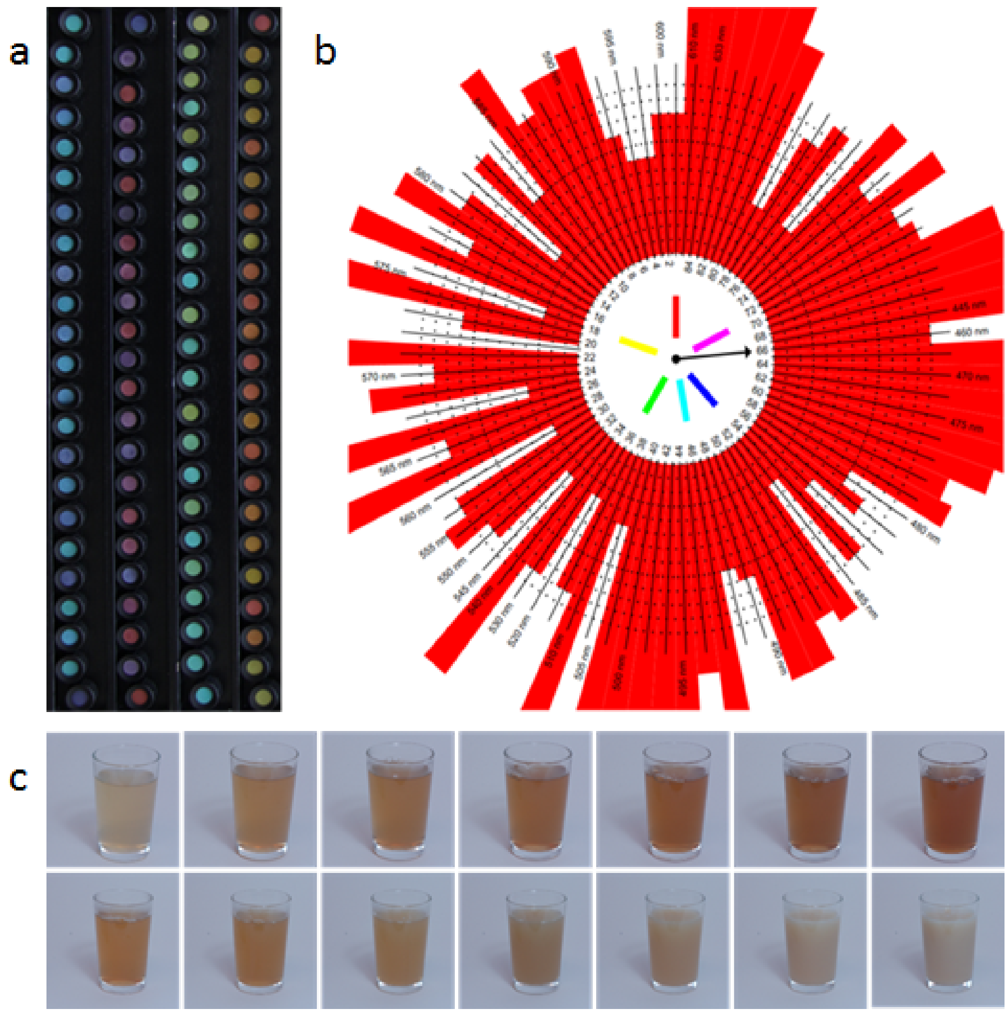

Figure 8.1: a) The Farnsworth-Munsell 100 Hue Test, as completed by MS. b) A plot of the results of MS's Farnsworth-Munsell test. c) Tea strength stimuli, ranging from weakest to strongest. d) Milk concentration stimuli, ranging from least to most milky.

8.3 Procedure: ranking of real stimuli

Two sets of photographs of real tea (details of creation can be found in the Supplementary Materials) varying in either tea strength or milkiness were used. Each set had seven levels of concentration of black tea or milk respectively (Figure 8.1c and d). On each trial, the observer was shown three stimuli (each about nine degrees of visual angle square) presented vertically in the centre of the screen (MS's hemianopia makes comparison of vertically arranged stimuli easier than horizontally arranged ones). The observer was asked to rank each set of three stimuli in order, from that with the lowest to the highest tea strength in one test session, or the lowest to highest milk concentration in another test session. For each task, all combinations of the seven stimuli were used, yielding 35 trials. MS requires additional time to complete trials and fatigues fairly quickly, so completed just one block of trials. Controls were asked to complete two blocks of trials.

8.4 Results

We scored performance by calculating the 'slope' of rankings on any trial. Regardless of the absolute values of strength we simply ordered the physical strengths 1, 2, and 3. We then regressed the order of judgements onto these. So if the reported order, 1, 2, 3, matches the physical order the slope would be 1.0, if the reported order was 3, 2, 1, the slope would be -1.0, if the reported order was 1, 3, 2 the slope would be 0.5 and so on. We then performed a single-sample t-test on the slopes of the ranking, where the null hypothesis (if judgements were random) was that the average slope would be zero. When ranking milk concentration, MS ranked the stimuli significantly differently from 0 (M = 0.26, SD = 0.623, t(34) = 2.44, p = 0.020) in the correct direction. When ranking tea strength, MS ordered the stimuli significantly differently from 0 (t(34) = -5.16, p<0.001) but in the 'wrong' direction (M = -0.53, SD = 0.61). All controls ranked the stimuli significantly differently from 0 for both tasks, in the correct direction (see Table 8.1 for t-test values). We then tested the difference between the single case mean and the sample mean, to determine whether MS's results were significantly different from those of the age-matched control and the two non-age-matched controls, using single-case methodology for a small sample (Crawford & Garthwaite, 2002). MS's results were significantly different from the controls for both the milkiness task (t(3) = -29.01, $p<0.001$) and the tea strength task (t(3) = -157.74, $p<.001$) - see Figure 8.2. The estimated percentage of the normal population falling below MS's score for the milkiness task was 0.059%, and 0.020% for the tea strength task.

Table 8.1: T-test results for each of the three control participants, on tea strength and milkiness ranking tasks.

Participant/task	t	df	Sig. (2-tailed)	Mean	Std. Deviation
HR / Milk	56.802	69	.000	0.9571	0.14098
HR / Tea	41.085	69	.000	0.9643	0.19637
YB / Milk	*	69	*	1.0000	0.00000
YB / Tea	45.667	69	.000	0.9786	0.17928
RC / Milk	80.269	69	.000	0.9786	0.10200
RC / Tea	45.667	69	.000	0.9786	0.17928

* this participant was correct on every trial, and so a t-test could not be conducted.

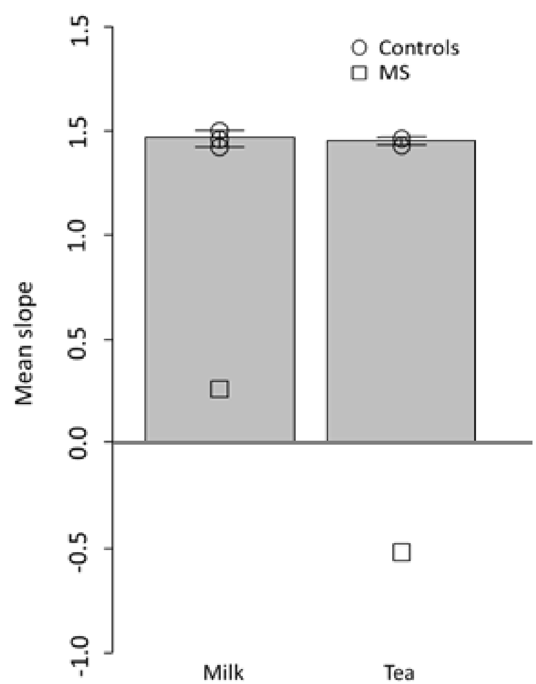

Figure 8.2: Bar chart showing the mean slope of rankings made by controls for milk and tea tasks with standard error bars, and individual means for controls and MS.

8.4.1 Image statistics

In order to investigate potential sources of information that might be used - by the controls or by MS - as pseudocues upon which to base decisions we ran a number of task simulations using image statistics of the tea-region of the images. We computed the mean, variance, and kurtosis of hue, saturation and value. We used a statistic from each image to simulate the ranking task. For a set of three stimuli, the simulated decision ranked the images according to ascending values of the statistic chosen. We then calculated the goodness of fit between these simulated responses and the data obtained from MS and each control participant. Mean saturation provided the best fit for the milkiness ranking task for all controls (mean squared error values for HR, YB, RC = 0.011, 0.000, and 0.005 respectively). Asymptote of saturation provided the same goodness of fit for this task as it essentially gave the same values as the mean saturation. For MS, the best fit for the milkiness task was variance of hue (mean squared error = 0.893), however this was still a poor fit - and also very unlikely to be an indication of the information used by MS, as MS is unable to discriminate hue. Mean value provided the best explanation for the responses on the tea ranking task by all controls (mean squared error values for HR, YB, RC = 0.020, 0.016, 0.016 respectively). For MS, the best fit for the tea task was provided by mean hue (mean squared error = 0.536), however again this was still a poor fit to the data. While the best fits of controls show that these observers seem to be making decisions in very similar ways, the best fits calculated for MS's responses were still poor and also based on hue, which MS is unable to discriminate. Therefore while mean saturation and mean value provide good explanations of the responses given by the controls (and might therefore be related to the pseudocues used by the observers) it is still not clear what MS might be basing his decisions on.

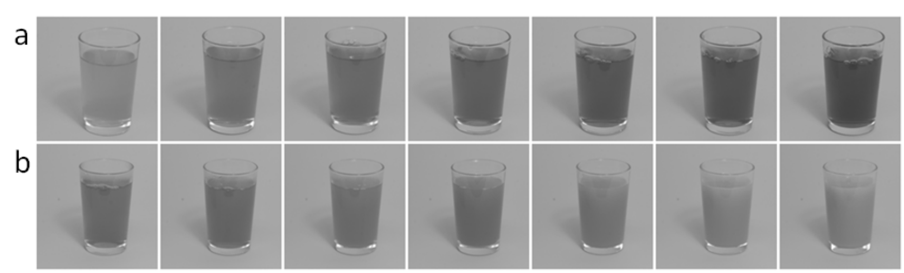

Figure 8.3: Greyscale versions of a) the stimuli varying in tea strength, and b) the stimuli varying in concentration of milk.

One obvious reason why MS might be so much poorer at ranking tea strength and milkiness is that he has no access to colour information. We therefore tested three additional normal controls on the same two ranking tasks, but using greyscale images (see Figure 8.3). All three observers were able to rank the stimuli well, with slopes that were significantly different from 0, on both tasks. Performance on the tea task was as good as that of controls tested with colour images (SK: M = 0.91, SD = 0.28, $t(69) = 27.13$, $p<0.001$, AC: M = 0.97, SD = 0.24, $t(69) = 34.00$, $p<0.001$, LN: M = 0.90, SD = 0.29, $t(69) = 25.97$, $p<0.001$) but performance was not as consistent for the milkiness ranking task (SK: M = 0.77, SD = 0.39, $t(69) = 16.67$, $p<0.001$, AC: M = 0.89, SD = 0.26, $t(69) = 28.75$, $p<0.001$, LN: M = 0.84, SD = 0.26, $t(69) = 20.09$, $p<0.001$). Goodness-of-fit was calculated for each of these observers to the simulated performances based on image statistics of value only: mean value provided the best fit for the tea task (mean squared errors = 0.136, 0.114, 0.143 for SK, AC and LN respectively), and kurtosis of value provided the best fit for the milkiness task (mean squared errors = 0.8, 1.029, 0.793), however both were poorer fits than the fits obtained with the controls tested with colour images. The goodness of fit of those statistics to MS's performance was comparable for the milkiness task (mean squared error = 1.021). The goodness of fit of the MS data on the tea strength task was considerably poorer than that of the greyscale controls (mean squared error = 0.579), and of course MS was ranking in the opposite order.

8.5 Discussion

MS makes systematic rankings of the stimuli for both milkiness and tea strength tasks. While MS's rankings were significantly different from chance, and he must therefore be capable of discriminating between the stimuli (most likely by using lightness), it appears that MS either had a different understanding of the criteria for rankings or was using different cues from those used by normal observers. On the tea strength task, MS ranked stimuli consistently, but in the opposite order to controls. It is possible that MS may simply be unable to access the absolute lightness values that may form the basis of his judgements and so, for example, ranked from 'lightest' to 'darkest' where controls would have ranked in the opposite direction. On the milkiness task, MS ranked the stimuli in the same direction as controls, although less consistently. MS did seem to understand the instructions he was given on both tasks. The fact that MS is able to make consistent judgements about aspects of translucence does not necessarily imply that he is doing so using the same mechanisms as normal observers.

The behavioural simulations based on image statistics showed that the statistic of best fit to control observers on the milkiness task was mean saturation. For the tea task, the best fit for all control observers was mean value (although a poor fit overall). We also calculated fits to the data from normal controls and MS with a model based on saturation gradients, and it did not provide a good fit in either case. The statistics of best fit for MS's performance were variance of hue and mean of hue respectively; however these were very poor fits and, as we have noted, are unlikely to be the basis of MS's decisions as he is unable to perceive colour. It is possible that MS's performance is simply a result of his achromatopsia, and perhaps normal observers would also perform poorly without colour information. We therefore tested the three additional controls with greyscale versions of the tasks.

Performance of controls with greyscale images was not qualitatively different from that with colour images. Controls that completed the task with the greyscale images were still able to rank tea strength as well as controls that were given colour images, but performance on the milkiness task - while slopes were significantly different from zero - was poorer when the images were greyscale. The observers completing the greyscale task reported that the milkiness task was much harder to complete in greyscale compared to the tea strength task, which reflected the difference in their performance on the two tasks. MS was also more consistent at ranking the tea strength stimuli - albeit while ranking them in the 'wrong' order - than he was at ranking milkiness, which he was still nevertheless able to do above chance. Colour information appeared to be important for normal observers in making judgements of milkiness. However, while colour might be an important cue when making judgements of translucence, the results of the greyscale controls show that some other cues to translucence perception do not rely on colour information. MS may therefore be able to exploit these cues. His performance is still poorer comparative to the greyscale controls. This may be due to an inability to process all of the non-chromatic cues that controls exploited. Alternatively, it may just be due to more general factors affecting his memory, attention and behaviour. In either case, we can conclude that while colour - and the regions responsible for its perception - is important for perception of translucence, there are other cues available processed in regions which are still intact in MS, and lead to a capacity to perceive translucence and discriminate its component properties.

8.6 Supplementary materials

8.6.1 Method of creating the real tea stimuli

Stimuli

To create controlled images of real tea, a 'master' tea solution was made with freshly boiled water before adjusting the strength by watering down as required. This volume was kept at a constant temperature to prevent the tannins precipitating and making the volume cloudy, as it was found that there was a just noticeable difference in the spectral composition of the light reflected from a glass of the tea solution detected by a spectroradiometer (Mahy et al., 1994) over a period of fifteen minutes as the liquid cooled. A procedure was developed to create the optimal strength of master solution. Ten teabags (Tetley) were added to three litres of boiled water for three minutes. The same brew of master solution was used for all stimuli. The total volume of liquid in each glass was 70ml, and the different stimuli were created by varying the amounts of tea solution, water, and milk (semi-skimmed, 1.8% fat).

An initial set of 20 stimuli was produced: 10 varying in milkiness, at the lowest strength of tea, and 10 varying in tea strength at the lowest level of milkiness. After piloting with multiple normal participants, an approximately linear perceptual scale was identified for each parameter within the range of stimuli, consisting of seven levels of milkiness while tea strength was held constant, and seven levels of tea strength while milk concentration was held constant.

Photographing the stimuli

Stimuli were photographed using a calibrated Nikon D80 camera which had been tested to ensure that no automatic lighting compensations were active in manual mode. Glasses of liquid were positioned against a white infinity-curve backdrop. The scene was lit by an overhead halogen lamp and a single fluorescent desk lamp with daylight spectrum, positioned to ensure some light passed through the volume of liquid towards the camera and also to create a visible shadow of the glass.

8.6.2 Apparatus

Experimental software was written in Matlab. Stimuli were presented on a calibrated NEC 2070SB CRT display (1064x768 pixels with refresh rate of 100Hz), controlled by means of a CRS Visage (Cambridge Research Systems). Observers were seated approximately 100cm in front of the screen in a blacked-out cubicle, and were asked to use a chin rest. Responses were made using a multi-button input device (Cedrus).

Chapter 9

General Discussion and Future Directions

9.1 Recent work on perceived gloss

Since the review of the perception of gloss in Chapter 2 was published, a number of additional studies within the field have been published. Research interest in this area has increased greatly in recent years, as outlined in the review, and is now moving quickly. A brief review of this new work is helpful in putting the thesis as a whole into the context of the current state of the field.

First, new industrially driven work has been undertaken setting out to devise a new metrological gloss scale, aiming to represent different levels of gloss, hue, roughness and refractive indices. This work primarily focused on measuring a small number of manufactured samples with different instruments to describe physical characteristics of the samples, and did not relate these measurements to perceptual judgements (Flys, Källberg, Ged, Silvestri, & Rosén, 2015).

Fores, Fairchild, and Tastl (2014) conducted a new investigation of Ferwerda et al.'s model of gloss perception (2001; previously discussed in Chapter 1.5) and determined that, outside the samples used to create the model, there was a perceptually non-uniform space among the dimensions. This is not surprising, given the fact that the model did not account for multiple aspects of perceived gloss, and was based on a limited set of impoverished stimuli. Fores et al. propose an alternative equation for the model, but again this does not account for all the cues that have been shown to be important for perceived gloss.

A variety of additional findings have been published, ranging across all stages from illumination to observer (as outlined in Chadwick and Kentridge 2015, Chapter 1.5). The complex spatial structure of highlights reflected from scene illumination was perceived as glossier than surfaces with simple highlight shapes produced by extended source illumination (van Assen, Wijntjes, & Pont, 2016), suggesting that constancy between simple and complex illuminations may be poor.

The general findings of Ho et al. (2008) that perceived gloss increased as surface 'bumpiness' increased were confirmed in a study by Qi, Chantler, Siebert, and Dong (2014). Increases in mesoscale roughness (or 'bumpiness') increased perceived gloss, but only up to a certain level of 'bumpiness' after which the surface was perceived to be rough again. The percentage of highlight area was found to correlate highly with perceived gloss.

The importance of inferring the 3D shape for perceiving material properties (including perceived gloss) was highlighted by Marlow and Anderson (2015). A number of different surface textures were applied to a pattern of luminance gradients (which was kept constant) to produce different three-dimensional shape interpretations. These stimuli appeared to show objects of different surface reflectances - shiny or matte - under different illuminations (front-lit and grazing angles, respectively). By varying texture and motion cues - both alone and in conjunction - they concluded that any cues that provide sufficient information about the 3D shape of a surface, such as motion and texture, can be used to interpret properties of the material. Specifically, the rate at which luminance varies across the surface with 3D shape provides a cue to gloss.

Wang, Pappas, and de Ridder (2015) explored new statistical cues using image transformations that increase perceived contrast. They explored this change in perceived contrast in relation to perceived gloss, and found that their transformation - a sub-band based S-curve transformation - increased perceived gloss. Contrast is of course already known to influence perceived gloss (Chapter 1.5, Hunter 1937). However Wang et al. suggest that the correlation between the standard deviation of their statistic for contrast and the perceived gloss of natural surfaces might serve as a source of information for the visual system. Whether this is the case is unclear, but they did find that this statistic was a better fit for perceived gloss than the skew of luminance, suggested by Motoyoshi et al. (2007).

In addition to research into purely visually-oriented material perception, there has also been an increase in research investigating how information from multiple senses might combine in decision-making. As described in Chapter 7, Adams et al. (2016) discovered that changes in friction affected perceived gloss. Changes in haptic friction produced a significant main effect with visual gloss on perceived gloss, and an interaction was also found between friction and visual gloss, where an increase in visual gloss and a decrease in 'slipperiness' counteracted one another. These wide-ranging approaches to understanding gloss perception contribute to a continually emerging picture of a multi-faceted perceptual experience dependent on a wide range of information in the scene. Combining many types of information would be computationally quite complex, and it is intuitive to conclude that processing of perceived gloss might only occur in the later stages of visual processing.

The neural correlates of colour and texture processing in humans and their place in the visual processing hierarchy have been investigated for some time; how does perceived gloss fit into the picture? It has already been suggested that the processing of gloss may not be totally reliant on areas responsible for colour or texture (Kentridge et al., 2012). Imaging studies and electrophysiological single-cell recording studies have chiefly investigated brain regions involved with the processing of gloss in macaque monkeys. There is broad agreement about the involvement of areas in the ventral visual cortex including the inferior temporal cortex (Komatsu, Nishio, Okazawa, & Goda, 2013; Nishio et al., 2012; Nishio, Shimokawa, Goda, & Komatsu, 2014; Okazawa, 2013; Okazawa et al., 2012a; Okazawa, Goda, & Komatsu, 2012b). More recently, fMRI studies have been undertaken with human observers, which largely corroborated these findings. Involvement of the posterior fusiform sulcus and V3B was identified, with gloss information apparently being processed differently by each area (Sun, Ban, Di Luca, & Welchman, 2015). The authors concluded that these mid-level areas may be particularly important in supporting perceived surface gloss. Furthermore, a number of studies discovered that there may be involvement from higher areas in processing perceived gloss in more complex circumstances. Motion flows of specular and matte surfaces indicated regions in higher cortex, but it was not clear how this was integrated with photometric cues (Kam, Mannion, Lee, Doerschner,

& Kersten, 2015). While ventral areas were already established as processing monocular cues to gloss, Sun, Di Luca, et al. (2016) found that higher dorsal areas were involved in binocular cues to gloss as well. They concluded that as transfer effects were found between the two there may be shared representations of cues between these areas. Similar differential processing was found for 2D and 3D cues for gloss (Sun, Di Luca, et al., 2015), and additional transfer effects were found from 2D to 3D cues between early V1/V2 and dorsal V3/V7.

9.2 This thesis

The work in this thesis was motivated by a lack of knowledge in the area of translucence perception, and the emerging need to investigate perceived gloss in the context of translucence. Work including that of Fleming and colleagues, and Anderson (Anderson, 2011; Fleming, 2014; Fleming, Jensen, & Bülthoff, 2004) highlights the importance of studying the perception of translucence and gloss in conjunction. The aims of this thesis were to investigate the perception of light scattering both in terms of translucence (scattering within volumes) and also in terms of the relation to physical and perceived gloss, as many objects appear shiny because of an upper glossy layer in the surface that allows some degree of light transport beneath the surface. To investigate these questions, a number of techniques were used, employing physically based ray tracing technology to explore how translucence and shine are perceived in multiple contexts and environments.

Before asking specific questions about translucence or gloss perception in specific contexts, it was important to establish precisely how these material properties are being interpreted by the visual system. As indicated in the introduction, previous work to date on the perception of translucence either made theoretical propositions about how the visual system could establish whether a material is translucent (where criteria might be necessary, but not sufficient) or took experimental approaches to manipulating translucence without testing observers. The study reported in Chapter 2 was needed to investigate precisely how observers may be estimating material properties. The approach was adapted specifically to the question, as it first allowed us to show that observers cannot be making estimates of the specific properties of light transport - absorption and scattering - as they were unable to completely 'perceptually un-mix' a liquid volume into its constituents when the components of the mixture were a pure scatterer and a pure absorber. On the other hand, they were relatively adept at perceptually un-mixing a real material, tea, in which the real materials being mixed - tea and milk - interact in scattering and absorption. The second experiment in Chapter 2, using simplified rendered stimuli, allowed us to test the hypothesis that observers were making shortcut estimates of light transport properties. Were they doing so, performance should have improved. If performance worsened, this would suggest that the cues used were more complex and not necessarily related to shortcut estimates of light transport properties. This study therefore provided the first experimental evidence for perceived translucence being driven by pseudocues (potentially complex spatially-related image statistics), which agreed with opinions in the literature (Fleming, 2014). While the conclusion itself is not different from the consensus proposal, no previous studies had addressed this experimentally. There had been no investigation of perceived translucence in relation to real light transport properties beyond simple image statistics. Similarly, no potential candidates had been proposed for complex statistics that could approximate the judgements of real observers.

As outlined earlier, such translucent materials are not always found in volumes - there are often thin layers of translucent materials near the surface of shiny objects. It was therefore important to undertake a similar experiment to that in Chapter 2, using similar techniques of manipulation of scatter and absorption in a layer (which also essentially manipulate physical gloss) to determine how similar judgements of materials were made in this context. Observers were asked to say which object (in a two alternative forced choice task) looked darker, or more dusty (on an opaque base material), as 'dusty/cloudy' equates to less specular reflectance from the surface, and so it appears less shiny. These changes were perceived in a similar way to those seen in volumes, in that judgements depended on a complex combination of the physical variables, and observers were not able to separate them completely.

The first two studies demonstrated that observers were making decisions of translucence or shine in relation to manipulations of scatter in volumes and layers, and that it is relatively easy to find statistical pseudocues that account for our results across conditions. Accordingly, an initial exploration of the constancy of these percepts could begin. How stable is our perception of the glossiness or translucence of a material under changes in viewing conditions? The first common contextual change we addressed was the importance of lighting direction. A recent study showed that there was a strong effect of lighting direction on perceived translucence (Xiao et al., 2014). Our first two studies showed that the method of combining techniques of MLDS and MLCM can robustly demonstrate how two physical manipulations contribute to observers' perceptions. The same method can be applied to assessment of constancy. We therefore used the method to test the contributions of lighting direction and light scattering within a material to perceived translucence. Our method of simulating scattering differed from that of Xiao et al., who manipulated the phase function of scattering - that is, the ratio of forward- and backward-scatter - whereas we varied the overall amount of scatter with equal amounts of forward- and backward-scattering. In volumetric materials our results showed perceived translucence was not necessarily determined by a straightforward contribution from the two variables. In materials where our manipulation of scattering was restricted to a thin subsurface layer the interaction between lighting direction and scattering material was, again, complex. In layers, observers found it easier to disentangle the two physical variables (light direction and physical scattering), suggesting that the contribution of light direction to perceived dustiness was less complex. This suggests that the pseudocues used by observers seem to serve us better when judging layers of coating rather than volumes. This is perhaps because the thinner layer allows more direct comparison of properties of the layer, as cues such as reflections or information visible behind the layer are available.

It also became apparent that it might be difficult to interpret these results in terms of perceived shine. Exploratory testing, and anecdotal comments from participants, suggested that the scattering layer did manipulate perceived gloss as predicted. Generally, perceived gloss decreased as physical scattering increased regardless of lighting direction, but front-lit objects looked shinier perhaps as there were more visible reflections in the surface of the object. It is also possible that light direction influenced perceived gloss as when objects were front lit specular highlights would be formed that were not present on the same back-lit objects. This supports the idea that, while physical gloss can be manipulated, the pseudocues employed by the visual system must involve factors beyond those directly affected by changes in physical gloss.

This is again evidence for pseudocues rather than inverse optics, and argues that the use of these pseudocues can vary greatly between contexts. The studies reported here, although not aiming to identify precisely the nature of the pseudocues employed by participants, suggest that observers seem to be using similar groups of pseudocues with which to make perceptual judgements. Individuals may weight these pseudocues slightly differently, producing slight variations in responses. In ambiguous scenes, they appear to try using different pseudocues to resolve the decision. The particular strategies of individuals may vary, perhaps depending on experience of previously encountered materials or on ability to discriminate variation in the pseudocues present, producing greater variation between participants on more difficult tasks.

So far, the studies summarised have established ways in which observers perceive characteristics of translucence, the ways in which this relates to perceived shine, and how perceptions of layers or volumes are interpreted under a simple contextual change in lighting direction. A fuller investigation of the way in which subsurface scattering layers affect perceived gloss must also address the way these layers interact with the base material they envelop. Is it the case that physically identical changes in a subsurface scattering layer affect the perceived gloss of objects equally, regardless of the base materials underlying the scattering layer? Chapter 5 filled this gap by comparing the results of three separate groups, each making decisions about the effects of subsurface scattering layers applied to a different base material. As reported in the previous chapter, increasing the scattering within a subsurface layer generally decreases perceived gloss. It was also the case that scattering and absorption did not have wholly independent effects on percepts of gloss or cloudiness and darkness of surfaces. The simplest additive interactions between physical absorption and scatter occurred on ceramic materials. On metal and pyrex materials the interactions were more complex, possibly as a result of the large variations in the reflected or transmitted images seen across the objects. Although such images provide cues to gloss through contrast or distinctness of images, they can be affected both by changes in scattering and changes in absorption within an enveloping layer. We can view these differences in the perceptual effects of identical changes in scattering or absorption of layers on different base materials as something akin to a failure of constancy. We see a great deal more agreement between participants than conflict, and while there are some individual differences the perceptual properties are generally judged in a similar way.

In the real world the surfaces of objects are often far from uniform. Surface layers that affect the glossiness of materials are likely to be subject to spatial variation. Subsurface scattering layers may not be uniform in thickness or may be subject to local damage. Material adhering to the surface of objects that affects glossiness such as dust, dirt or grease is likely to have been acquired randomly, and also rubbed off randomly. To understand how we perceive the glossiness of real world objects we need to investigate the effects of deviation from spatial uniformity of scattering. How do we interpret irregularities and different levels of variation in changes in physical gloss? Chapter 6 attempted to explore this question by manipulating layers of scattering materials on opaque objects. It showed that thickness of the layer (or the amount of scattering in a layer) affected perceived gloss as one might expect. The degree of spatial variation in the layer also affected perceived gloss, over and above the effect of the absolute amount of scattering. As the degree of scatter becomes spatially less uniform across a surface, it is perceived as progressively less glossy.

This study also allowed us to question the nature of decision making at a cognitive level - that is, whether asking observers a slightly different question in relation to reaching judgements might induce a top-down effect on the way in which decisions were made about the same set of stimuli. We asked one group of participants to make judgements about glossiness (a visual property) and another group to make judgements about cleanliness, a more abstract property, albeit one that might be related to glossiness. High levels of physical gloss could be akin to judging a smooth surface to be free from imperfections such as dirt - but do we judge shine and cleanliness in the same way? The weightings of the two physical variables were significantly different for observers making judgements of shine compared to the weightings of observers making judgements of perceived cleanliness. This provides evidence for a model of perceptual decision making where changes at multiple levels of information between object and observer - physical, and interactions of physical materials; perception, and interactions in perceptions; and contributions at a cognitive level - can all affect the decisions made by observers. It shows that, in conjunction with results from the previous studies, these different levels can each change the priorities of the weighted pseudocues available in a scene, both in cognitive and pre-cognitive processes.

The studies so far have investigated how a range of different contextual factors contribute to our decision making; but what happens when observers receive conflicting information about a surface from different senses? Recent studies have found that tactile information such as friction influence perceived gloss, and have even discovered an interaction in perceived gloss with friction (Adams et al., 2016). But there are other kinds of tactile information - tactile roughness was isolated as one of the principal components important for identifying material properties and material categorisation (Baumgartner et al., 2013). Adams et al. (2016) manipulated 'slipperiness' (friction) with a Phantom (a haptic simulation device), but not fine scale roughness. We set out to determine whether fine scale roughness (probably transduced by different kinds of mechanoreceptor in the skin from friction) had a similar influence over perceived gloss. While there was some effect of roughness on perceived gloss, there was no overall interaction, and considerable variation between observers in how they combined the cues. Three of eight observers did demonstrate an interaction in how they combined the cues, but the remaining five did not. This is a potentially interesting and important finding, as it suggests that information received by different mechanoreceptors might be combined in different ways - and also be weighted or prioritised by individuals in different ways, depending on how important they judge that information. An adaptation experiment comparing different forms of tactile information may provide some insight into how these kinds of sensory information are combined when making decisions about material properties. Participants would be asked to repeatedly touch a material of a given surface roughness (with the aim of adapting them to that level of surface roughness), and would then be asked to make visually mediated roughness judgements of surfaces varying in roughness while simultaneously touching the tactile stimulus. Adaptation effects on judgements of roughness - depending on the level of roughness used in the adapting stimulus - would suggest that perceptual processing of tactile roughness influenced visual judgements of surface texture. This task could be repeated with friction, to test whether these kinds of tactile feedback are being combined differently with visual judgements in perceptual estimates.

To reiterate, the results reported thus far have provided evidence in favour of a theory of vision where observers are using pseudocues to interpret translucence and gloss, potentially involving complex spatial-related image statistics. But how is processing of these pseudocues related to the processes underlying perception of other material properties? Is processing of perceived translucence reliant on areas of the visual cortex responsible for processing other material properties? Recent studies have shown that colour and texture are processed independently, although information about both is brought into conjunction in some brain areas (Cavina-Pratesi et al., 2010a, 2010b). Kentridge et al. (2012) also showed that the processing of perceived gloss was performed by areas distinct from those involved in colour processing. How might translucence relate to these other material properties? We tested this hypothesis by asking MS, a cerebral achromatopsic with a range of other visual deficits, to make judgements of translucence and darkness of liquid volumes (the stimuli created for part 1 of Chapter 2). MS was able to make distinctions between liquids of different translucence, but less able to make judgements of darkness. This reflected his poor ability to differentiate different levels of lightness, but demonstrated that he was able to detect differences in translucence that were separate from changes in absorption. This suggests that decisions of translucence are not necessarily based on lightness alone. It also provides evidence for the cortical basis of the processing of translucence being at least partially separate from the areas responsible for processing colour.

However, this raises further questions concerning the processing of gloss, texture, and translucence. MS was able to detect differences in both tea strength and milkiness, so it is possible that the processing of these properties is separate, or that there might be more shared processing for these properties. Research into cortical processing of gloss has recently become more prominent, in both humans and macaques. The inferior temporal cortex in particular has been proposed as a region involved in selectively responding to different types of physical gloss, and representing different kinds of perceived gloss (Komatsu et al., 2013; Nishio et al., 2012). The superior temporal sulcus has also been implicated (Nishio et al., 2014; Okazawa, 2013; Okazawa et al., 2012a, 2012b), and higher regions of the visual pathway have been proposed as well. These higher regions may be involved in representing 3D cues to gloss, and differences in monocular and binocular cues to gloss (posterior fusiform sulcus, V3B/KO, Sun, Ban, et al. 2015; Sun, Di Luca, et al. 2016, 2015). Yet there is little information about the areas involved in perceived translucence. There is a clear need for evidence from fMRI studies concerning the regions involved in perceptual judgements of translucence.

9.3 Further research: an fMRI experiment

The findings of this thesis raise a number of new questions. An experimental approach to the question of how different kinds of tactile information might be combined has already been proposed. What regions of cortex are responsible for the processing of translucence - might they be related to those involved in perceived gloss? Here I propose a novel experiment, applying recently developed psychophysical techniques, which aims to address this.

9.3.1 Investigating the neural correlates of the perception of translucence using a new application of the maximum likelihood conjoint measurement (MLCM) method of analysis

To date, many studies have investigated the potential neural correlates of the perception of various object and material properties, including perceived size (Cavina-Pratesi, Goodale, & Culham, 2007; Op de Beeck, Torfs, & Wagemans, 2008; Sperandio, Chouinard, & Goodale, 2012), 3D surface structures (Taira, Nose, Inoue, & Tsutsui, 2001), shape (Kourtzi & Kanwisher, 2001), colour (Beauchamp, Haxby, Jennings, & DeYoe, 1999) and texture (Beason-Held et al., 1998). It has also been recently shown that processing of different perceptual properties such as colour and texture are mediated by separate cortical areas but come together in others (Cavina-Pratesi et al., 2010a, 2010b). Here I discuss the design for an experiment to investigate the neural loci involved in translucence perception. Chapter 2 reports an experiment which is the first systematic experimental investigation of potential ways that physical aspects of translucence are interpreted by the human visual system. Besides the more general conclusions outlined in that paper, this experiment produced behavioural data which illustrated the way in which participants made judgements of translucence. In that experiment variation of milk concentration in tea essentially manipulates translucence of the liquid (Chapter 2), and in the proposed neuroimaging study, that task would form the behavioural basis of the experiment. These behavioural data in conjunction with functional magnetic resonance imaging would potentially be able to help identify regions in the brain whose patterns of response mimicked perceptual judgements.

9.3.2 Method

Stimuli

The stimuli used in this experiment would be images of real cups of tea, varying in the strength of the tea and in milk concentration, the production of which is detailed in the supplemental information section of Chapter 2. There are approximately equal perceptual steps between levels of these stimuli, as a result of a difference scaling procedure.

Procedure

In the scanner, participants would be asked to complete a simple two alternative forced choice task. This task would essentially be the same conjoint measurement task as that outlined in Chapter 2 - where two images of cups of tea, varying in both tea strength and milk concentration, are presented - with the exception that stimuli would be presented one after the other, with a small delay between presentation times. Participants would be asked to make judgements of relative milk concentration, regardless of the strength of tea, and would indicate whether they thought the first or second image presented had a higher concentration of milk regardless of the strength of the tea in each volume by pressing one of two buttons. After a small washout period, the next trial would begin. 720 trials in total would be presented, in 6 repeated blocks of 120; the task would be completed over the course of 3-4 one hour sessions in the scanner. The actual judgements made by participants would also be recorded and analysed with maximum likelihood conjoint measurement analysis.

Participants

For a pilot study the participants could be two of the authors of the paper in Chapter 2. This would not pose an issue for the task itself or the results, as it is a purely perceptually-based judgement. It would have the advantage that the participant would be familiar with the task. Both potential participants have normal or corrected-to-normal vision.

Data analysis

The data obtained would initially be processed as normal within SPM, however instead of using analysis of variance we would apply maximum likelihood conjoint measurement analysis techniques of BOLD signal strength on a voxel-by-voxel basis (of regions of visual cortex) in an attempt to ascertain which regions were involved in the decision making process.

The usual analysis in SPM involves collating the BOLD signal strength values into matrices, with different matrices for each time point. These data would then be processed within SPM using GLM to produce, for example, analyses of variance between experimental conditions. Instead of performing this final analysis, we would extract the matrices of values and process these independently outside SPM. Each voxel would be treated separately, and the higher activation above baseline between the two stimuli for each trial would be treated as a 'decision' in favour of a higher perceived milk concentration. The 'decisions' of each voxel would then be processed independently using maximum likelihood conjoint measurement analysis (Knoblauch & Maloney, 2012). This analysis would produce three log likelihoods, from fitting the decisions to the three hypothetical models (independent, additive and saturated), from which a model of best fit could be determined. It would also provide the model itself for each set of decisions, showing how 'decisions' of milkiness are related to the two manipulated variables (tea strength and milk concentration). These multiple models of best fit would then be compared to the actual obtained behavioural data - of both the judgements made by the participants while in the scanner, and the data obtained in Chapter 2 - to ascertain whether any voxels showed a similar pattern of response to the task.

This method has advantages beyond the usual method of analysis of variance. Analysis of variance could identify potential neural regions of involvement in the perception of translucence, but it would not be able to show where the particular cues that drive behaviour are being computed. Using maximum likelihood conjoint measurement in conjunction with behavioural data would enable us to investigate more specifically the potential neural regions involved in making comparative judgements of translucence based on those fundamental cues. That is, any regions identified as having similar patterns of response to the behavioural data could, rather than just being involved in some lower level of processing that leads to translucence perception, be more specifically identified in coding perceived translucence itself.

This technique, if proven viable, would have potential application in many other fields as a means of identifying areas coding the values of varying perceptual properties that influence decisions.

9.4 Further research: a multidimensional approach

It is clear that a wide range of variables influence perceived gloss, including the translucence properties of subsurface scattering layers. Previous attempts to characterise the complex nature of multiple cues to gloss have often downplayed multidimensionality. The various manipulations that have been used in exploring cues to gloss also range from simple image editing to complex contextual changes in the environment. An exploration of the multidimensionality of perceived gloss must separate different classes of variation in scenes. I therefore suggest that multidimensional analysis of cue contributions is necessary in order to understand perceived gloss.

9.4.1 The perception of gloss - a constellation of pseudocues: unpicked

Extensive recent research in the perception of gloss has revealed that rather than estimating physical dimensions, or employing short-cut statistical calculations, the visual system makes judgements of gloss based on a large group of factors - a constellation of pseudocues (Chadwick & Kentridge, 2015; Fleming, 2014). The contribution of these factors varies based on situation, information available in any given context, and also by observer. Weightings and rankings vary a great deal from one choice to the next. In order to try and unpick the contributions of the multiple dimensions of perceived gloss, we propose an experiment aiming to identify the ways in which the many factors involved in perceived gloss contribute to our judgements, while attempting to minimise the level of variability between observers.

It is clear from the large body of work investigating perceived gloss that perceptual judgements are determined by many factors. Several attempts to characterise the visual decision making process have been made to date (Billmeyer & O'Donnell, 1987; Ferwerda et al., 2001; Motoyoshi et al., 2007), but each might be criticised for omitting factors shown elsewhere to be important for making consistent and informed decisions (Chadwick & Kentridge, 2015). For example, Billmeyer and O'Donnell perform multidimensional scaling on sets of glossy painted sheets, but this was performed on individual sets of stimuli varying in one of Hunter's types of gloss alone. These stimuli could, therefore, not capture possible interactions between these types. Unidimensional solutions were found for each subset of stimuli, which also required different solutions for different levels of luminance, suggesting that unidimensionality is not necessarily the best approximation for the results. Anderson (2013) describes a similar viewpoint: experiments with stimuli limited by 'laboratory conditions' might not explain how variables relate to observers' usual interpretations in normal (un-impoverished) circumstances. I therefore propose that a more wide-ranging approach to the multidimensional scaling technique might better serve the questions arising within the field and begin to provide better indications of the underlying method of the visual system in making judgements of perceived gloss.

9.4.2 Method

General method

Since we are interested in how observers utilise different sources of information in making perceptual judgements, and how the different sources contribute towards the decision, multi-dimensional scaling (MDS) is an appropriate method to use as it allows the comparison of many objects and calculates solutions ranging from 1-6 dimensions. A comparison method also allows these judgements to be made without causing problems of vocabulary or arbitrary rankings or ratings. It would be interesting to see if individuals varied in how they weight or rank the various factors; there may still be wide variation even if more information is available than from limited and impoverished stimuli.

Variables

A large range of variables have been demonstrated to have influence over perceived gloss (Chadwick & Kentridge, 2015), such as illumination type and direction, the lightness, shape and texture of the surface of an object, and the presence of binocular or motion cues. The inclusion of as many different types of gloss, and factors involved in perceived gloss, would enable a more realistic interpretation of the task to be achieved by observers. It would also aim to reduce the extent to which observers are forced to make decisions based on information lacking key qualities that would normally contribute to their judgements, which potentially limits the generality and ecological validity of results. Controlling for - or excluding - a large number of factors already established as influencing perceived gloss and only allowing decisions to be made on limited manipulated parameters could potentially produce complex interactions of the two factors that might not be present were other information available for use and comparison, thereby skewing the representation of the decision making process.

In standard investigations of how properties are perceived, the number of manipulated parameters is limited to one or two, while all others are excluded or controlled, to ensure that the results obtained are a valid reflection of judgements made on those parameters alone. While this method is scientifically rigorous, it has limitations. Real materials rarely vary to such small degrees in only a limited number of physical parameters, and a great deal more information is available to observers which they can use to further differentiate perceived properties. It is possible that limiting the number of informative variables in a scene may produce more simplistic relationships or more complex interactions of the variables that would not have otherwise been found were naturalistic information available, thereby skewing the representation of the decision-making process - and subsequently, our characterisation of it.

The large number of variables which have been shown to influence perceived gloss can be divided into two categories - real scene manipulations, and manipulations of images. Real scene manipulations - such as a changes in illumination, the shape of the object, surface lightness, and viewing distance - are factors that vary in real life which affect perceived gloss, whereas image manipulations constitute factors that have been found to affect perceived gloss but which are obtained by manipulating the image itself. These sorts of image manipulations most generally affect Hunter's six types of gloss (sheen, haze, distinctness of image, absence of surface texture, specular highlights, and contrast - Hunter 1937). While it would be of interest to investigate how these types of gloss influence each other in judgements of perceived gloss, it does not make sense to compare them alongside real world manipulations, as combinations of these changes are not encountered by observers. Therefore, in performing multidimensional scaling experiments, we propose performing two separate experiments, one for each type of variation.

Stimuli

For the experiment with image manipulations, a full set of stimuli would be generated realistically using physically based rendering software, varying in the presence or absence of Hunter's six types of gloss (totalling 64 stimuli). For real life manipulations, the four most prominent factors affecting perceived gloss would be used to generate a second set of realistic stimuli (broadband illumination versus limited artificial illumination, object shape, surface lightness, and viewing distance), while Hunter's six types of gloss would all be present (totalling 16 stimuli). Manipulation of these factors would be achieved using realistic PBRT-based ray tracers, and automated with Matlab. There are a significant number of other variables which affect perceived gloss - motion in particular - but these would not be technically feasible with current equipment. To overcome objections to a lack of stereoscopically presented stimuli, an eyepatch could be used, or each of the stimuli could be rendered twice (from slightly different perspectives) and presented using a stereoscope.

Procedure

The task itself would involve presenting observers with a range of these stimuli simultaneously, one of which would be designated the anchor stimulus, and observers would be asked to rank the remaining stimuli presented in order of most similar to the anchor stimulus (known as the conditional rank ordering task - Schiffman, Young, and Reynolds 1981, or alternatively the anchor stimulus method, Borg and Groenen 2005, Giguère 2006). Once the anchor stimulus has been presented and removed, the stimulus chosen as most similar would then be removed and given the highest ranking, and the next most similar stimulus would be chosen, and so on. Once all stimuli in that trial have been ranked, a new set of images would then be presented with a different anchor stimulus, and the task repeated. This experiment would therefore be using ordinal levels, with stimuli arranged in order of magnitude, but only comparative ranking information would be available. A precise difference between values would always be the same for measurement levels.

It is certain that 64 stimuli would be far too many to present to an observer at any one time, and would be impossible to rank. Even 16 stimuli may be too many. It is, nevertheless, still possible to perform the task in such a way as to obtain a good set of data with which to perform multidimensional scaling analysis and to ensure the task is achievable by observers (Giguère, 2006; Tsogo, Masson, & Bardot, 2000). From the full number of stimuli, n, a subset of size k would be selected, where k would optimally be 8 or 9, as the maximum number of items that participants can usually rank at once is approximately 9 (Wilson & Sharples, 2015). One stimulus would be selected as the anchor stimulus, and separated from the rest - all other stimuli would then be ranked, from most to least similar to the anchor stimulus. Descriptions of the conditional rank ordering task stipulate that once ranked, stimuli are removed from the set - however Wilson and Sharples (2015) point out that ranking of similar objects, particularly when the differences are small, is less accurate than when all stimuli can be seen at once and moved around. Therefore all k stimuli would be presented on screen simultaneously, and participants would be allowed to move the stimuli around to record their rankings in relation to the anchor stimulus. This would be repeated n-1 times, rotating the object used as the standard anchor. While this would not produce a full comparison of all stimuli with all possible standard anchors, this method has been shown to provide an accurate picture of relationships among variables. If necessary, multiple blocks could be performed in concurrent sessions of testing, with alternative subsets of k selected, and the results combined, or the results from multiple participants combined in an average matrix.

Results

The format of the results themselves would be as follows: for each participant, an nxn matrix would be produced, with each row and column representing one of the n objects. Diagonal elements would be set to 1, and each row would be constructed on the basis of conditional rank ordering, where the standard anchor object would be the row object. Entries corresponding to the k ranked column objects would contain integers from 2 to k+1, according to the column object rank, where 2 indicated the object ranked as most similar, and k+1 the object ranked as least similar. The remaining (n-k-1) entries would be assigned an average rank of ((n+k+2)/2). If using multiple blocks, multiple matrices from each individual would be averaged. Matrices from all individuals could, if desired, be averaged, to obtain a group scaling solution using non-metric scaling. Kruskal (1964) recommends transforming similarity values into dissimilarity values for use with MDS, as the relationship between perceived psychological distance and dissimilarities is direct and positive (Giguère, 2006; Kruskal, 1964). The matrix values can be transformed by subtracting the original data values from a constant which is higher than all scores collected (i.e. $(k{+}1){+}1$ or more).

This set of matrices could then be used as input for MDS calculations to be performed by such a program such as SPSS or R. Non-metric MDS (used for ordinal ranking) essentially generates random configurations of points, calculates distances between them and finds the optimal monotonic transformation of them to obtain the closest solution for those points, and then increases the goodness of fit by finding a new configuration based on these calculations. The analysis performs 1-6 dimensional solutions, and indicates the number of dimensions of the most appropriate solution, although in SPSS it is often better to compute each possible solution individually and then compare them. The stress function (a particular kind of goodness-of-fit) is used to compare to some criterion, to determine how good the solution is. R-squared is also calculated, to show the proportion of variance which can be accounted for within the optimally scaled data by the MDS analysis (R-squared of 0.6 considered the minimum level, with 0.9 used for non-metric analysis). The validity of an MDS solution is - for non-metric MDS - generally given by the value of the optimised stress value (critical values for evaluating a scaling solution using a stress value are given by Kruskal, 1964b, as seen in Giguère 2006), although some have argued that this is not always the best criterion by which to judge the best MDS solution (Tsogo et al., 2000). Therefore it would be better to use both the stress function and the R-squared value.

From the results obtained, we would arrive at a solution of best fit of x dimensions. The distances provided by the analysis between the different stimuli, in the x dimensions, would provide us with the means to determine the factors of importance in making perceptual judgements of gloss, and how the variables affect one another. Most importantly, these results would help to inform us of the physical dimensions along which the perceptual gloss space varies. In the case of the experiment manipulating Hunter's types of gloss, the results would offer the possibility of experimentally validating the six types as important to observers when making judgements.

9.5 Summary

By employing novel, recently developed psychophysical methods, physically accurate computer rendered images, manipulations of real stimuli, and cross-modal approaches, we have shown the range of phenomena in human observers' perceptual experience of translucence and gloss - two relatively unexplored material properties. We have explored the relationship between physical and perceptual spaces of translucence and gloss in a variety of contexts, and have begun to address and characterise the kinds of information that are used by observers when making decisions about translucence and gloss. Our results support the conclusion that these sources of information for perceived translucence and gloss are likely to be composed of pseudocues, which are capable of characterising complex material properties, but are not directly based on the physical attributes of the material such as estimates of the physical light transport properties. These pseudocues may be based on complex spatially-related image statistics, but additional factors would need to be accounted for in producing estimates of those properties that allowed the perceptual experience to be consistent across contextual changes. The image statistics we have identified would not necessarily be invariant in the face of those changes (such as object shape). However, it is also true that our perceptions of translucence and gloss may only be weak in constancy for changes in some contexts.

We have used a method for manipulating gloss by adjusting scattering in a subsurface layer of coating, and shown it affects the perceived gloss of a material. This method should be of broad interest, since it allows identical manipulations of gloss to be applied to objects with different surface characteristics and shapes. We found that varying physical gloss using this method affected perceived gloss differently when the layer was applied to different base materials, and also when the homogeneity of the layer varied. Real layers found in materials have much more complex layering than the layers simulated here, however, and many other studies could be undertaken involving realistic models of the translucence of, for example, human skin.

Developing our understanding of translucence and gloss perception has a variety of potential applications. The appearance of plastics, coatings and paints in industry could be optimised, as the perceived appearance of these materials significantly influences the perceived quality of products made using these materials. Industries are also currently basing much of their research on measuring physical gloss without accounting for perceptual types of gloss when developing new paints and coatings, and so knowledge of how observers perceive gloss and translucence may improve the efficiency of this work. Realistic models of translucence in relation to human skin would be beneficial in a number of fields. Medical prosthetics could be made to look more realistic, and understanding the appearance of skin in terms of translucent layers could inform broader applications in cosmetics. Recent changes to the types of materials allowed in cosmetic products (such as the ban on micro-beads and micro-particles) mean that new solutions are required for existing applications - changing appearances without using micro-particles or other ingredients could well be achieved by using layers of translucent substances.

This thesis has by no means solved all of the issues in the perception of gloss and translucence. There are other more general questions - for example, concerning particular pseudocues for gloss and translucence that are invariant to contextual changes, and also questions about how we perceive translucence of more complex natural layered materials such as skin. We have only begun to scratch the surface of all of the questions that might be asked. What we have done is to show that it is possible to understand various components of translucence and gloss perception, but there is much remaining to be explored.

References

Adams, W. J. (2016). The development of audio-visual integration for temporal judgements. *PLoS Computational Biology*, *12*(4), 1004865.

Adams, W. J., & Elder, J. H. (2014). Effects of specular highlights on perceived surface convexity. *PLoS computational biology*, *10*(5), e1003576.

Adams, W. J., Kerrigan, I. S., & Graf, E. W. (2016). Touch influences perceived gloss. *Scientific reports*, *6*.

Adelson, E. H. (2001). On seeing stuff: The perception of materials by humans and machines. In *Photonics west 2001-electronic imaging* (p. 1-12). International Society for Optics and Photonics.

Adelson, E. H., & Anandan, P. (1990). *Ordinal characteristics of transparency.* Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology.

Aernouts, B., Van Beers, R., Watté, R., Huybrechts, T., Lammertyn, J., & Saeys, W. (2015). Visible and near-infrared bulk optical properties of raw milk. *Journal of dairy science*, *98*(10), 6727-6738.

Anderson, B. L. (2011). Visual perception of materials and surfaces. *Current Biology*, *21*(24), R978-R983. doi: http://dx.doi.org/10.1016/j.cub.2011.11.022

Anderson, B. L. (2013). *The perceptual representation of transparency, lightness, and gloss.* Handbook of perceptual organization. Oxford: Oxford University Press.

Anderson, B. L., & Kim, J. (2009). Image statistics do not explain the perception of gloss and lightness. *Journal of Vision*, *9*(11).

Arend, L. E., Reeves, A., Schrillo, J., & Goldstein, R. (1991). Simultaneous color constancy: papers with diverse munsell values. *J. Opt. Soc. Am A*, *8*(4), 661–672.

Baumgartner, E., Wiebel, C. B., & Gegenfurtner, K. R. (2013). Visual and haptic representations of material properties. *Multisensory research*, *26*(5), 429-455.

Baumgartner, E., Wiebel, C. B., & Gegenfurtner, K. R. (2015). A comparison of haptic material perception in blind and sighted individuals. *Vision research*, *115*, 238-245.

Beason-Held, L. L., Purpura, K. P., Krasuski, J. S., Maisog, J. M., Daly, E. M., Mangot, D. J., ... VanMeter, J. W. (1998). Cortical regions involved in visual texture perception: a fmri study. *Cognitive Brain Research*, *7*(2), 111-118.

Beauchamp, M. S., Haxby, J. V., Jennings, J. E., & DeYoe, E. A. (1999). An fmri version of the farnsworth–munsell 100-hue test reveals multiple color-selective areas in human ventral occipitotemporal cortex. *Cerebral Cortex*, *9*(3), 257-263.

Beck, J. (1964). The effect of gloss on perceived lightness. *The American Journal of Psychology*, *77*(1), 54-63. doi: 10.2307/1419271

Beck, J., Prazdny, K., & Ivry, R. (1984). The perception of transparency with achromatic colors. *Perception & Psychophysics*, *35*(5), 407-422.

Beck, J., & Prazdny, S. (1981). Highlights and the perception of glossiness. *Attention, Perception, & Psychophysics*, *30*(4), 407-410.

Beekman, F. J., den Harder, J. M., Viergever, M. A., & van Rijk, P. P. (1997). Spect scatter modelling in non-uniform attenuating objects. *Physics in medicine and biology*, *42*(6), 1133.

Bergmann Tiest, W. M., & Kappers, A. M. L. (2007). Haptic and visual perception of roughness. *Acta Psychologica*, *124*(2), 177-189.

Berzhanskaya, J., Swaminathan, G., Beck, J., & Mingolla, E. (2005). Remote effects of highlights on gloss perception. *Perception-London*, *34*(5), 565-576.

Billmeyer, F. W., & O'Donnell, F. X. D. (1987). Visual gloss scaling and multidimensional scaling analysis of painted specimens. *Color Research & Application*, *12*(6), 315-326. doi: http://dx.doi.org/10.1002/col.5080120606

Blake, A., & Bülthoff, H. (1990). Does the brain know the physics of specular reflection? *Nature*, *343*(6254), 165-168.

Boes, D. C., Graybill, F. A., & Mood, A. M. (1974). *Introduction to the theory of statistics.* McGraw-Hill, New York.

Borg, I., & Groenen, P. J. (2005). *Modern multidimensional scaling: Theory and applications.* Springer Science & Business Media.

Bouvier, S. E., & Engel, S. A. (2006). Behavioral deficits and cortical damage loci in cerebral achromatopsia. *Cerebral Cortex*, *16*(2), 183-191.

Brill, M. H. (1984). Physical and informational constraints on the perception of transparency and translucency. *Computer Vision, Graphics, and Image Processing*, *28*(3), 356-362.

Burgess, A. (1985). Visual signal detection. iii. on bayesian use of prior knowledge and cross correlation. *OSA A*, *2*(9), 1498-1507.

Cant, J. S., & Goodale, M. A. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cerebral cortex*, *17*(3), 713-731.

Cavina-Pratesi, C., Goodale, M. A., & Culham, J. C. (2007). Fmri reveals a dissociation between grasping and perceiving the size of real 3d objects. *PLoS ONE*, *2*(5), e424. doi: 10.1371/journal.pone.0000424

Cavina-Pratesi, C., Kentridge, R. W., Heywood, C. A., & Milner, A. D. (2010a). Separate channels for processing form, texture, and color: evidence from fmri adaptation and visual object agnosia. *Cerebral cortex*, *20*(10), 2319-2332.

Cavina-Pratesi, C., Kentridge, R. W., Heywood, C. A., & Milner, A. D. (2010b). Separate processing of texture and form in the ventral stream: evidence from fmri and visual agnosia. *Cerebral cortex*, *20*(2), 433-446.

Chadwick, A. C., & Kentridge, R. W. (2015). The perception of gloss: A review. *Vision research*, *109*, 221-235.

Crawford, J. R., & Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia*, *40*(8), 1196-1208.

Cubitt, S., Palmer, D., & Tkacz, N. (2015). *Digital light.* Open Humanities Press.

Di Luca, M. (2016, July). *Multisensory material properties: how do the senses tell us how soft an object is?* Seminar presented at the Department of Psychology, University of Durham.

Doerschner, K., Boyaci, H., & Maloney, L. T. (2010). Estimating the glossiness transfer function induced by illumination change and testing its transitivity. *Journal of Vision*, *10*(4). doi: 10.1167/10.4.8

Doerschner, K., Fleming, R. W., Yilmaz, O., Schrater, P. R., Hartung, B., & Kersten, D. (2011). Visual motion and the perception of surface material. *Current Biology*, *21*(23), 2010-2016.

Donner, C., Weyrich, T., d'Eon, E., Ramamoorthi, R., & Rusinkiewicz, S. (2008). A layered, heterogeneous reflectance model for acquiring and rendering human skin. *ACM Transactions on Graphics (TOG)*, *27*(5), 140.

Dror, R. O., Adelson, E. H., & Willsky, A. S. (2001). Estimating surface reflectance properties from images under unknown illumination. In *SPIE Photonics West: Human vision and electronic imaging VI* (p. pp. 231-242). (10.1117/12.429494)

Dror, R. O., Willsky, A. S., & Adelson, E. H. (2004). Statistical characterization of real-world illumination. *Journal of Vision*, *4*(9). doi: 10.1167/4.9.11

Ernst, M. O., & Banks, M. S. (2002). Using visual and haptic information for discriminating objects. *Journal of Vision*, *2*(10), 126-126. (10.1167/2.10.126) doi: http://dx.doi.org/10.1167/2.10.126

Ernst, M. O., Banks, M. S., & Bülthoff, H. H. (2000). Touch can change visual slant perception. *Nature neuroscience*, *3*(1), 69-73.

Ferwerda, J. A., Pellacini, F., & Greenberg, D. P. (2001). Psychophysically based model of surface gloss perception. In *Photonics West 2001-electronic imaging* (p. 291-301). International Society for Optics and Photonics.

Fleming, R. W. (2014). Visual perception of materials and their properties. *Vision Research*, *94*(0), 62-75. doi: http://dx.doi.org/10.1016/j.visres.2013.11.004

Fleming, R. W., & Bülthoff, H. H. (2005). Low-level image cues in the perception of translucent materials. *ACM Transactions on Applied Perception (TAP)*, *2*(3), 346-382.

Fleming, R. W., Dror, R. O., & Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, *3*(5).

Fleming, R. W., Jäkel, F., & Maloney, L. T. (2011). Visual perception of thick transparent materials. *Psychological Science*, *22*(6), 812-820.

Fleming, R. W., Jensen, H. W., & Bülthoff, H. H. (2004). Perceiving translucent materials. In *Proceedings of the 1st symposium on applied perception in graphics and visualization* (p. 127-134). ACM.

Fleming, R. W., Torralba, A., & Adelson, E. H. (2004). Specular reflections and the perception of shape. *Journal of Vision*, *4*(9). doi: 10.1167/4.9.10

Fleming, R. W., Wiebel, C., & Gegenfurtner, K. (2013). Perceptual qualities and material classes. *Journal of Vision*, *13*(8). doi: 10.1167/13.8.9

Flys, O., Källberg, S., Ged, G., Silvestri, Z., & Rosén, B. (2015). Characterization of surface topography of a newly developed metrological gloss scale. *Surface Topography: Metrology and Properties*, *3*(4), 045001.

Fores, A., Fairchild, M. D., & Tastl, I. (2014). *Perceptual gloss space BRDF projection, uniformity validation, and lightness distance metric.* ACM. doi: 10.1145/2628257.2628355

Formankiewicz, M. A., & Mollon, J. (2009). The psychophysics of detecting binocular discrepancies of luminance. *Vision Research*, *49*(15), 1929-1938.

Foster, D. H., & Nascimento, S. M. (1994). Relational colour constancy from invariant cone-excitation ratios. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *257*(1349), 115-121.

Ged, G., Obein, G., Silvestri, Z., Le Rohellec, J., & Viénot, F. (2010). Recognizing real materials from their glossy appearance. *Journal of Vision*, *10*(9), 18.

Giguère, G. (2006). Collecting and analyzing data in multidimensional scaling experiments: A guide for psychologists using spss. *Tutorials in Quantitative Methods for Psychology*, *2*(1), 27-38.

Gkioulekas, I., Xiao, B., Zhao, S., Adelson, E. H., Zickler, T., & Bala, K. (2013). Understanding the role of phase function in translucent appearance. *ACM Transactions on Graphics (TOG)*, *32*(5), 147.

Hanada, M. (2012). Difference between highlight and object colors enhances glossiness. *Perceptual and motor skills*, *114*(3), 735-747.

Harrison, V., & Poulter, S. (1951). Gloss measurement of papers-the effect of luminance factor. *British Journal of Applied Physics*, *2*(4), 92.

Hartung, B., & Kersten, D. (2002). Distinguishing shiny from matte. *Journal of Vision*, *2*(7), 551-551.

Hasni, I., Bourassa, P., Hamdani, S., Samson, G., Carpentier, R., & Tajmir-Riahi, H.-A. (2011). Interaction of milk -and -caseins with tea polyphenols. *Food Chemistry*, *126*(2), 630-639.

Hetherington, M. J., Martin, A., MacDougall, D. B., Langley, K. R., & Bratchell, N.

(1990). Comparison of optical and physical measurements with sensory assessment of the ripeness of tomato fruit lycopersicon esculentum. *Food quality and preference*, *2*(4), 243-254.

Heywood, C. A., Cowey, A., & Newcombe, F. (1991). Chromatic discrimination in a cortically colour blind observer. *European Journal of Neuroscience*, *3*(8), 802-812.

Heywood, C. A., & Kentridge, R. W. (2003). Achromatopsia, color vision, and cortex. *Neurologic clinics*, *21*(2), 483-500.

Ho, Y.-X., Landy, M. S., & Maloney, L. T. (2006). How direction of illumination affects visually perceived surface roughness. *Journal of Vision*, *6*(5).

Ho, Y.-X., Landy, M. S., & Maloney, L. T. (2008). Conjoint measurement of gloss and surface texture. *Psychological Science*, *19*(2), 196-204.

Ho, Y.-X., Maloney, L. T., & Landy, M. S. (2007). The effect of viewpoint on perceived visual roughness. *Journal of Vision*, *7*(1).

Hunter, R. (1937). Methods of determining gloss. *Journal of Research of the National Bureau of Standards*, *18*(1), 19-41.

Hurlbert, A., Cumming, B., & Parker, A. (1991). Recognition and perceptual use of specular reflections. In *Investigative ophthalmology & visual science* (Vol. 32, p. 1278-1278). LIPPINCOTT-RAVEN PUBL 227 EAST WASHINGTON SQ, PHILADELPHIA, PA 19106.

Ingersoll, L. R. (1921). The glarimeter an instrument for measuring the gloss of paper. *Journal of the Optical Society of America*, *5*(3), 213-215. doi: 10.1364/josa.5.000213

Jacquemoud, S., & Ustin, L. S. (2008). Modeling leaf optical properties. *Photobiological Sciences Online*.

Jensen, H. W., Marschner, S. R., Levoy, M., & Hanrahan, P. (2001). A practical model for subsurface light transport. In *Proceedings of the 28th annual conference on computer graphics and interactive techniques* (p. 511-518). ACM.

Ji, W., Pointer, M. R., Luo, R. M., & Dakin, J. (2006). Gloss as an aspect of the measurement of appearance. *JOSA A*, *23*(1), 22-33.

Johansson, R. S., & Flanagan, J. R. (2009). Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nature Reviews Neuroscience*, *10*(5), 345-359.

Kam, T.-E., Mannion, D. J., Lee, S.-W., Doerschner, K., & Kersten, D. J. (2015). Human visual cortical responses to specular and matte motion flows. *Frontiers in human neuroscience*, *9*.

Keane, T. J. (1989). U.s. patent no. 4,886,355.

Kentridge, R. W., Heywood, C. A., & Cowey, A. (2004). Chromatic edges, surfaces and constancies in cerebral achromatopsia. *Neuropsychologia*, *42*(6), 821-830.

Kentridge, R. W., Thomson, R., & Heywood, C. A. (2012). Glossiness perception can be mediated independently of cortical processing of colour or texture. *cortex*, *48*(9), 1244-1246.

Kerrigan, I. S., & Adams, W. J. (2013). Highlights, disparity, and perceived gloss with convex and concave surfaces. *Journal of Vision*, *13*(1). doi: 10.1167/13.1.9

Kerrigan, I. S., Adams, W. J., & Graf, E. W. (2010). Does it feel shiny? haptic cues affect perceived gloss. *Journal of Vision*, *10*(7), 868-868.

Kersten, D. (2000). 25 high-level vision as statistical inference. *The new cognitive neurosciences*, 353.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as bayesian inference. *Annu. Rev. Psychol.*, *55*, 271-304.

Kim, J., & Anderson, B. L. (2010). Image statistics and the perception of surface gloss and lightness. *Journal of Vision*, *10*(9).

Kim, J., Marlow, P., & Anderson, B. L. (2011). The perception of gloss depends on highlight congruence with surface shading. *Journal of Vision*, *11*(9).

Kim, J., Marlow, P. J., & Anderson, B. L. (2012). The dark side of gloss. *Nature Neuroscience*, *15*(11), 1590-1595.

Kinnear, P., & Sahraie, A. (2002). New farnsworth-munsell 100 hue test norms of normal observers for each year of age 5–22 and for age decades 30–70. *British Journal of Ophthalmology*, *86*(12), 1408-1411.

Knoblauch, K., & Maloney, L. T. (2012). *Modeling psychophysical data in R* (Vol. 32). Springer.

Koenderink, J., & Pont, S. (2003). The secret of velvety skin. *Machine vision and applications*, *14*(4), 260-268.

Koenderink, J. J., & van Doorn, A. J. (1980). Photometric invariants related to solid shape. *Journal of Modern Optics*, *27*(7), 981-996.

Koenderink, J. J., & van Doorn, A. J. (2001). Shading in the case of translucent objects. In *Proc. SPIE 4299, human vision and electronic imaging VI, 312* (Vol. 4299, p. 312-320). (10.1117/12.429502)

Komatsu, H., Nishio, A., Okazawa, G., & Goda, N. (2013). 'yellow'or 'gold'?: Neural processing of gloss information. In *Computational color imaging* (p. 1-12). Springer.

Körding, B. U.-M. W. J. Q. S. T. J. B. . S. L., K. P. (2007). Causal inference in multisensory perception. *PLoS one*, *2*(9), 943.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, *293*(5534), 1506-1509. doi: 10.1126/science.1061133

Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences*, *96*(1), 307-312. doi: 10.1073/pnas.96.1.307

Krol, . E.-D. W., M. E. (2011). When believing is seeing: The role of predictions in shaping visual perception. *The Quarterly Journal of Experimental Psychology*, *64*(9), 1743-1771.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1-27.

Landy, M. S. (2007). Visual perception: A gloss on surface properties. *Nature*, *447*(7141), 158-159. (10.1038/nature05714)

Lavin, M. A. (1973). The gloss of glossy things. *MIT Artificial Intelligence Laboratory Working Papers, WP-41, Vision Flash*(No. 41).

Leloup, F. B., Obein, G., Pointer, M. R., & Hanselaer, P. (2013). Toward the soft metrology of surface gloss: A review. *Color Research & Application*, n/a-n/a. doi: 10.1002/col.21846

Leloup, F. B., Pointer, M. R., Dutré, P., & Hanselaer, P. (2010). Geometry of illumination, luminance contrast, and gloss perception. *JOSA A*, *27*(9), 2046-2054.

Leloup, F. B., Pointer, M. R., Dutré, P., & Hanselaer, P. (2012). Overall gloss evaluation in the presence of multiple cues to surface glossiness. *JOSA A*, *29*(6), 1105-1114.

Lichtenauer, M. S., Schuetz, P., & Zolliker, P. (2013). Interaction improves perception of gloss. *Journal of Vision*, *13*(14), 14.

Lindstrand, M. (2005). Instrumental gloss characterization - in the light of visual evaluation: A review. *Journal of Imaging Science and Technology*, *49*(1), 61-70.

Mahy, M., Eycken, L., & Oosterlinck, A. (1994). Evaluation of uniform color spaces developed after the adoption of cielab and cieluv. *Color Research & Application*, *19*(2), 105-121.

Maloney, L. T., & Yang, J. N. (2003). Maximum likelihood difference scaling. *Journal of Vision*, *3*(8), 5-5.

Marlow, P., Kim, J., & Anderson, B. L. (2011). The role of brightness and orientation congruence in the perception of surface gloss. *Journal of Vision*, *11*(9).

Marlow, P. J., & Anderson, B. L. (2013). Generative constraints on image cues for perceived gloss. *Journal of Vision*, *13*(14), 2.

Marlow, P. J., & Anderson, B. L. (2015). Material properties derived from three-dimensional shape representations. *Vision Research*, *115, Part B*, 199-208.

Marlow, P. J., Kim, J., & Anderson, B. L. (2012). The perception and misperception of specular surface reflectance. *Current Biology*, *22*(20), 1909-1913.

Metelli, F. (1970). An algebraic development of the theory of perceptual transparency. *Ergonomics*, *13*(1), 59–66.

Metelli, F. (1974a). Achromatic color conditions in the perception of transparency. In R. B. MacLeod & H. L. Pick (Eds.), *Perception: Essays in honor of James J. Gibson* (p. 317). Ithaca, NY, US: Cornell University Press.

Metelli, F. (1974b). The perception of transparency. *Scientific American*, *230*(4), 90–98.

Methven, T. S., & Chantler, M. J. (2012). Problems of perceiving gloss on complex surfaces. In *Proceedings of the 3rd international conference on appearance* (p. p. 43-47). Lulu Press.

Middleton, W. E. K., & Mungall, A. G. (1952). The luminous directional reflectance of snow. *Journal of the Optical Society of America*, *42*(8), 572-579.

Mollon, J., Newcombe, F., Polden, P., & Ratcliff, G. (1980). On the presence of three cone mechanisms in a case of total achromatopsia. *Colour vision deficiencies*, *5*, 130-135.

Mori, M. K. F., M., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98-100.

Motoyoshi, I. (2010). Highlight–shading relationship as a cue for the perception of translucent and transparent materials. *Journal of vision*, *10*(9), 6-6.

Motoyoshi, I., & Matoba, H. (2012). Variability in constancy of the perceived surface reflectance across different illumination statistics. *Vision Research*, *53*(1), 30-39.

Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature*, *447*(7141), 206-209.

Murakoshi, T., Masuda, T., Utsumi, K., Tsubota, K., & Wada, Y. (2013). Glossiness and perishable food quality: visual freshness judgment of fish eyes based on luminance distribution. *PloS one*, *8*(3), e58994.

Muryy, A. A., Welchman, A. E., Blake, A., & Fleming, R. W. (2013). Specular reflections and the estimation of shape from binocular disparity. *Proceedings of the National Academy of Sciences*, *110*(6), 2413-2418. doi: 10.1073/pnas.1212417110

Narasimhan, S. G., Gupta, M., Donner, C., Ramamoorthi, R., Nayar, S. K., & Jensen, H. W. (2006). Acquiring scattering properties of participating media by dilution. *ACM Transactions on Graphics (TOG)*, *25*(3), 1003-1012.

Nefs, H. T., Koenderink, J. J., & Kappers, A. M. L. (2006). Shape-from-shading for matte and glossy objects. *Acta Psychologica*, *121*(3), 297-316.

Nishida, S., Motoyoshi, I., & Maruya, K. (2011). Luminance-color interactions in surface gloss perception. *Journal of Vision*, *11*(11), 397. doi: 10.1167/11.11.397

Nishida, S., Motoyoshi, I., Nakano, L., Li, Y., Sharan, L., & Adelson, E. (2008). Do colored highlights look like highlights? *Journal of Vision*, *8*(6), 339.

Nishida, S., & Shinya, M. (1998). Use of image-based information in judgments of surface-reflectance properties. *JOSA A*, *15*(12), 2951-2965.

Nishio, A., Goda, N., & Komatsu, H. (2012). Neural selectivity and representation of gloss in the monkey inferior temporal cortex. *The Journal of Neuroscience*, *32*(31), 10780-10793.

Nishio, A., Shimokawa, T., Goda, N., & Komatsu, H. (2014). Perceptual gloss parameters are encoded by population responses in the monkey inferior temporal cortex. *The Journal of Neuroscience*, *34*(33), 11143-11151.

Obein, G., Knoblauch, K., & Viénot, F. (2004). Difference scaling of gloss: Nonlinearity, binocularity, and constancy. *Journal of Vision*, *4*(9).

O'Donnell, F. X., & Billmeyer, F. W. (1986). Psychometric scaling of gloss. *Review and Evaluation of Appearance: Methods and Techniques. ASTM STP*, *914*, 14-32.

Okazawa, G. (2013). Localization of regions activated by surface gloss in macaque visual cortex using fMRI.

Okazawa, G., Goda, N., & Komatsu, H. (2012a). P1-27: Localizing regions activated by surface gloss in macaque visual cortex by fMRI. *i-Perception*, *3*(9), 641-641.

Okazawa, G., Goda, N., & Komatsu, H. (2012b). Selective responses to specular surfaces in the macaque visual cortex revealed by fMRI. *NeuroImage*, *63*(3), 1321-1333.

Olkkonen, M., & Brainard, D. H. (2010). Perceived glossiness and lightness under real-world illumination. *Journal of Vision*, *10*(9). doi: 10.1167/10.9.5

Olkkonen, M., & Brainard, D. H. (2011). Joint effects of illumination geometry and object shape in the perception of surface reflectance. *i-Perception*, *2*(9), 1014.

Op de Beeck, H. P., Torfs, K., & Wagemans, J. (2008). Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. *The Journal of Neuroscience*, *28*(40), 10111-10123. doi: 10.1523/jneurosci.2511-08.2008

Pfund, A. (1930). The measurement of gloss. *J. Opt. Soc. Am*, *20*, 23-26.

Pharr, M., & Humphreys, G. (2004). *Physically based rendering: From theory to implementation*. Morgan Kaufmann.

Pizlo, Z. (2001). Perception viewed as an inverse problem. *Vision Research*, *41*(24), 3145-3161. doi: http://dx.doi.org/10.1016/S0042-6989(01)00173-0

Podrebarac, S. K., Goodale, M. A., & Snow, J. C. (2014). Are visual texture-selective areas recruited during haptic texture discrimination? *NeuroImage*, *94*, 129-137.

Pont, S. C., & te Pas, S. F. (2006). Material–illumination ambiguities and the perception of solid objects. *Perception*, *35*(10), 1331.

Purves, D., Augustine, G., Fitzpatrick, D., Hall, W., LaMantia, A., McNamara, J., & White, L. (2008). Neuroscience. *De Boeck, Sinauer, Sunderland, Mass*, Mechanoreceptors Specialized to Receive Tactile Information. Available from: http://www.ncbi.nlm.nih.gov/books/NBK10895/.

Qi, L., Chantler, M. J., Siebert, J. P., & Dong, J. (2012). How mesoscale and microscale roughness affect perceived gloss. In *Predicting perceptions: Proceedings of the 3rd international conference on appearance* (p. p 48-51). Lulu Press, Inc.

Qi, L., Chantler, M. J., Siebert, J. P., & Dong, J. (2014). Why do rough surfaces appear glossy? *Journal of the Optical Society of America A*, *31*(5), 935-943.

Ramachandran, V. (1985). The neurobiology of perception. *Perception*, *14*(2), 97–103.

Sakano, Y., & Ando, H. (2008). Effects of self-motion on gloss perception. *Perception*, *37*, 77.

Schiffman, S. S., Young, F. W., & Reynolds, M. L. (1981). *Introduction to multidimensional scaling: Theory, methods, and applications.* New York: Academic Press.

Serikawa, S., & Shimomura, T. (1993). Method for measuring glossiness of plane surfaces based on psychological sensory scale. *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences*, *76*(3), 439-446.

Sève, R. (1993). Problems connected with the concept of gloss. *Color Research & Application*, *18*(4), 241-252. doi: 10.1002/col.5080180407

Sharan, L., Rosenholtz, R., & Adelson, E. (2009). Material perception: What can you see in a brief glance? *Journal of Vision*, *9*(8), 784. doi: 10.1167/9.8.784

Singh, M., & Anderson, B. L. (2002a). Perceptual assignment of opacity to translucent surfaces: The role of image blur. *Perception*, *31*(5), 531-552.

Singh, M., & Anderson, B. L. (2002b). Toward a perceptual theory of transparency. *Psychological review*, *109*(3), 492.

Singh, M., & Hoffman, D. D. (1998). Part boundaries alter the perception of transparency. *Psychological Science*, *9*(5), 370-378.

Sperandio, I., Chouinard, P. A., & Goodale, M. A. (2012). Retinotopic activity in v1 reflects the perceived and not the retinal size of an afterimage. *Nature Neuroscience*, *15*(4), 540-542. (10.1038/nn.3069)

Sun, H.-C., Ban, H., Di Luca, M., & Welchman, A. E. (2015). fmri evidence for areas that process surface gloss in the human visual cortex. *Vision research*, *109*, 149-157.

Sun, H.-C., Di Luca, M., Ban, H., Muryy, A., Fleming, R. W., & Welchman, A. E. (2016). Differential processing of binocular and monocular gloss cues in human visual cortex. *Journal of neurophysiology*, *115*(6), 2779-2790.

Sun, H.-C., Di Luca, M., Fleming, R., Muryy, A., Ban, H., & Welchman, A. (2015). Brain processing of gloss information with 2d and 3d depth cues. *Journal of vision*, *15*(12), 818-818.

Sun, H.-C., Welchman, A. E., Chang, D. H., & Di Luca, M. (2016). Look but don't touch: Visual cues to surface structure drive somatosensory cortex. *NeuroImage*, *128*, 353–361.

Taira, M., Nose, I., Inoue, K., & Tsutsui, K. I. (2001). Cortical areas related to attention to 3D surface structures based on shading: An fMRI study. *NeuroImage*, *14*(5), 959-966.

te Pas, S. F., & Pont, S. C. (2005). A comparison of material and illumination discrimination performance for real rough, real smooth and computer generated smooth spheres. In *Proceedings of the 2nd symposium on applied perception in graphics and visualization* (p. 75-81). ACM.

Todd, J. T., Norman, J. F., & Mingolla, E. (2004). Lightness constancy in the presence of specular highlights. *Psychological Science*, *15*(1), 33-39.

Tsogo, L., Masson, M., & Bardot, A. (2000). Multidimensional scaling methods for many-object sets: A review. *Multivariate Behavioral Research*, *35*(3), 307-319.

van Assen, J. J. R., Wijntjes, M. W. A., & Pont, S. C. (2016). Highlight shapes and perception of gloss for real and photographed objects. *Journal of Vision*, *16*(6), 6.

Vandenbroucke, A., Fahrenfort, J., Meuwese, J., Scholte, H., & Lamme, V. (2016). Prior knowledge about objects determines neural color representation in human visual cortex. *Cerebral cortex*, *26*(4), 1401-1408.

Vangorp, P., Laurijssen, J., & Dutré, P. (2007). The influence of shape on the perception of material reflectance. In *ACM transactions on graphics (TOG)* (Vol. 26, p. 77). ACM.

Victor, J. D. (1988). Evaluation of poor performance and asymmetry in the farnsworth-munsell 100-hue test. *Investigative ophthalmology & visual science*, *29*(3), 476-481.

Viénot, F. (2012). Gloss characterization: a cyclic approach. In *Predicting perceptions: Proceedings of the 3rd international conference on appearance* (pp. 7–9). Edinburgh, UK: Lulu Press.

Wada, A., Sakano, Y., & Ando, H. (2014). Human cortical areas involved in perception of surface glossiness. *NeuroImage*, *98*, 243–257.

Wang, J., Pappas, T. N., & de Ridder, H. (2015). Effects of contrast adjustment on visual gloss of natural textures. In *SPIE/IS&T electronic imaging* (Vol. 9394, pp.

93940F–93940F). (10.1117/12.2078256)

Wendt, G., Faul, F., Ekroll, V., & Mausfeld, R. (2010). Disparity, motion, and color information improve gloss constancy performance. *Journal of Vision*, *10*(9).

Wendt, G., Faul, F., & Mausfeld, R. (2008). Highlight disparity contributes to the authenticity and strength of perceived glossiness. *Journal of Vision*, *8*(1), 14.

Wijntjes, M. W. A., & Pont, S. C. (2010). Illusory gloss on lambertian surfaces. *Journal of Vision*, *10*(9). doi: 10.1167/10.9.13

Wilson, J. R., & Sharples, S. (2015). *Evaluation of human work.* CRC Press.

Xiao, B., & Brainard, D. H. (2008). Surface gloss and color perception of 3d objects. *Visual neuroscience*, *25*(3), 371.

Xiao, B., Walter, B., Gkioulekas, I., Zickler, T., Adelson, E., & Bala, K. (2014). Looking against the light: How perception of translucency depends on lighting direction. *Journal of vision*, *14*(3), 17.